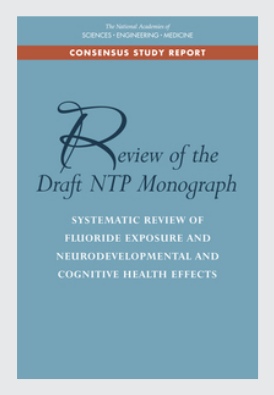

Review of the Draft NTP Monograph: Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects (2020)

DETAILS

60 pages | 6 x 9 | PAPERBACK

ISBN 978-0-309-67313-6 | DOI 10.17226/25715

CONTRIBUTORS

Committee to Review the NTP Monograph on the Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects; Board on Environmental Studies and Toxicology; Division on Earth and Life Studies; National Academies of Sciences, Engineering, and Medicine

SUGGESTED CITATION

National Academies of Sciences, Engineering, and Medicine 2020. *Review of the Draft NTP Monograph: Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects.* Washington, DC: The National Academies Press. https://doi.org/10.17226/25715.

Review of the Draft NTP Monograph

SYSTEMATIC REVIEW OF FLUORIDE EXPOSURE AND NEURODEVELOPMENTAL AND COGNITIVE HEALTH EFFECTS

Committee to Review the NTP Monograph on the Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects

Board on Environmental Studies and Toxicology

Division on Earth and Life Studies

A Consensus Study Report of

The National Academies of

SCIENCES • ENGINEERING • MEDICINE

THE NATIONAL ACADEMIES PRESS

Washington, DC

www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

This project was supported by Contract HHSN263201800029I between the National Academies of Sciences, Engineering, and Medicine and the Department of Health and Human Services. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the organizations or agencies that provided support for this project.

International Standard Book Number-13: 978-0-309-67313-6
International Standard Book Number-10: 0-309-67313-5
Digital Object Identifier: https:// doi.org/10.17226/25715

Additional copies of this publication are available from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.

Printed in the United States of America

Suggested citation: National Academies of Sciences, Engineering, and Medicine. 2020. *Review of the Draft NTP Monograph: Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects*. Washington, DC: The National Academies Press. https:// doi.org/10.17226/25715.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. John L. Anderson is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at **www.nationalacademies.org**.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Consensus Study Reports published by the National Academies of Sciences, Engineering, and Medicine document the evidence-based consensus on the study's statement of task by an authoring committee of experts. Reports typically include findings, conclusions, and recommendations based on information gathered by the committee and the committee's deliberations. Each report has been subjected to a rigorous and independent peer-review process and it represents the position of the National Academies on the statement of task.

Proceedings published by the National Academies of Sciences, Engineering, and Medicine chronicle the presentations and discussions at a workshop, symposium, or other event convened by the National Academies. The statements and opinions contained in proceedings are those of the participants and are not endorsed by other participants, the planning committee, or the National Academies.

For information about other products and activities of the National Academies, please visit www.nationalacademies.org/about/whatwedo.

COMMITTEE TO REVIEW THE NTP MONOGRAPH ON THE SYSTEMATIC REVIEW OF FLUORIDE EXPOSURE AND NEURODEVELOPMENTAL AND COGNITIVE HEALTH EFFECTS

Members

DAVID A. SAVITZ *(Chair)*, Brown University School of Public Heath, Providence, RI
KEVIN M. CROFTON, R3Fellows, Durham, NC
AKHGAR GHASSABIAN, New York University School of Medicine, New York, NY
JUDITH B. KLOTZ, Drexel University Dornsife School of Public Health, Philadelphia, PA
JULEEN LAM, California State University, East Bay, Hayward, CA
PAMELA J. LEIN, University of California, Davis, CA
GERMAINE M. BUCK LOUIS, George Mason University, Fairfax, VA
MICHAEL L. PENNELL, The Ohio State University, Columbus, OH
CRAIG STEINMAUS, University of California, Berkeley, Oakland, CA
CHARLES V. VORHEES, University of Cincinnati College of Medicine, Cincinnati, OH
KIMBERLY YOLTON, Cincinnati Children's Hospital Medical Center and University of Cincinnati College of Medicine, Cincinnati, OH

Staff

ELLEN K. MANTUS, Project Director
SUSAN MARTEL, Senior Program Officer
RADIAH ROSE-CRAWFORD, Manager, Editorial Projects
SARAH HARPER, Senior Program Assistant

Sponsor

NATIONAL INSTITUTE OF ENVIRONMENTAL HEALTH SCIENCES

BOARD ON ENVIRONMENTAL STUDIES AND TOXICOLOGY

Members

WILLIAM H. FARLAND (*Chair*), Colorado State University, Fort Collins, CO
LESA AYLWARD, Summit Toxicology, LLP, Falls Church, VA
ANN M. BARTUSKA, Resources for the Future, Washington, DC
GERMAINE M. BUCK LOUIS, George Mason University, Fairfax, VA
E. WILLIAM COLGLAZIER, American Association for the Advancement of Science, Washington, DC
FRANCESCA DOMINICI, Harvard University, Boston, MA
GEORGE GRAY, The George Washington University, Washington, DC
R. JEFFREY LEWIS, ExxonMobil Biomedical Sciences, Inc., Annandale, NJ
LINSEY C. MARR, VA Polytechnic Institute and State University, Blacksburg, VA
R. CRAIG POSTLEWAITE, Department of Defense, Burke, VA
REZA J. RASOULPOUR, Corteva Agriscience, Indianapolis, IN
IVAN RUSYN, Texas A&M University, College Station, TX
DEBORAH L. SWACKHAMER, University of Minnesota, St. Paul, MN
JOSHUA TEWKSBURY, Future Earth, Boulder, CO
SACOBY M. WILSON, University of Maryland College Park, MD

Staff

CLIFFORD S. DUKE, Director
RAYMOND A. WASSEL, Scholar and Director of Environmental Studies
LAURA LLANOS, Finance Business Partner
TAMARA DAWSON, Program Associate

Acknowledgments

This consensus study report was reviewed in draft form by persons chosen for their diverse perspectives and technical expertise. The purpose of this independent review is to provide candid and critical comments that will assist the National Academies of Sciences, Engineering, and Medicine in making each published report as sound as possible and to ensure that it meets institutional standards of quality, objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

We thank the following for their review of this report:

Ana Navas-Acien, Columbia University
David Bellinger, Harvard Medical School
Weihsueh Chiu, Texas A&M University
David Dorman, North Carolina State University
Mary Gilbert, US Environmental Protection Agency
Jayanth Kumar, California Department of Public Health
Karen Robinson, Johns Hopkins University

Although the reviewers listed above provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations of this report, nor did they see the final draft before its release. The review of the report was overseen by Martin Philbert, University of Michigan, and Jonathan Samet, Colorado School of Public Heath, who were responsible for making certain that an independent examination of the report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content rests entirely with the authoring committee and the National Academies.

The committee gratefully acknowledges the staff of the National Institute of Environmental Health Sciences for their presentations to the committee during open sessions. The committee is also grateful for the assistance of Norman Grossblatt who served as the report editor.

Trial Exhibit 653.009

Contents

APPENDIX

Summary

Over the last few decades, people have debated the benefits and hazards of fluoride exposure. The Centers for Disease Control and Prevention has recognized water fluoridation as one of the greatest public-health achievements, and others have claimed that fluoride exposure causes various adverse health effects. In 2006, the National Academies of Sciences, Engineering, and Medicine (the National Academies) reviewed the scientific literature on the health effects of fluoride exposure and concluded that chronic fluoride exposure can cause enamel fluorosis and weakening of bone that could increase the risk of fracture. Studies of the potential neurotoxicity of fluoride exposure lacked sufficient detail and did not allow definitive conclusions. However, the National Academies report concluded that the consistency of the results on neurotoxicity warranted further investigation. Since the 2006 report, several epidemiologic studies of fluoride exposure and neurodevelopmental and cognitive effects have been published. That research and a nomination from the Fluoride Action Network (FAN) prompted the National Toxicology Program (NTP) to conduct a systematic review of the evidence of adverse neurodevelopmental and cognitive effects of fluoride exposure. NTP's conclusions are summarized in the monograph *Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects.*[1]

To ensure the integrity of its report, NTP asked the National Academies to review the monograph. As a result of that request, the National Academies convened the present committee. It is important to note that the committee was tasked with reviewing the monograph and focused its efforts on evaluating whether evidence as presented in the monograph supported NTP's conclusions. Thus, it did not conduct its own independent evaluation of the evidence, and it did not conduct a data audit (an independent review of all the data reported in the monograph to identify omissions or errors in reporting). However, it did review some key literature to enable its review of the monograph. The committee's findings and suggestions for improvements are contained in this report; some overarching findings concerning methods, assessment of animal and human evidence, and NTP's hazard conclusion are provided here.

METHODS AND COMMUNICATION

The protocol for the systematic review described in the monograph was published on NTP's Web site in June 2017 and made available for public com-

[1]Referred to hereafter as the monograph.

ment. It was reviewed several times by technical advisers selected for their expertise on this topic. In general, it describes the overall systematic-review process and clearly outlines modifications that were made during the review. Thus, NTP appears to be adhering to best practices for systematic reviews with respect to the availability and documentation of such a protocol before initiation of a review.

The committee, however, identified several issues associated with the protocol. First, the role of the Office of Health Assessment and Translation (OHAT) handbook in developing the protocol is unclear. The protocol scarcely refers to the OHAT handbook and does not discuss the role of the handbook in its development. That ambiguity leads to concerns about the lack of detail in the protocol and about apparent conflicts between methodologic approaches in the protocol and the handbook itself.

Second, important details are missing from the protocol, including information on the strategy used to update the experimental animal literature, expertise and experience of review team members, and the planned conduct of statistical analyses. It does not provide explicit exclusion and inclusion criteria for study selection, which are critical for transparency of the process and reproducibility of the findings. It also does not provide justification for some of its decisions, for example, regarding screening parameters or what information to make publicly available, such as the list of excluded studies.

Third, there are some inconsistencies in the details provided in the protocol and the methods ultimately implemented in the monograph, including how mechanistic data would be considered, how the outcome assessment would be conducted, and which confounders were identified as critical covariates. Those discrepancies are troubling because inconsistencies between the protocol and the monograph raise questions about how the process was actually conducted, about what changes were made, and about when and why modifications were implemented.

The committee found some issues associated with data presentation and communication of various aspects of the process that are discussed further in the context of the evaluation of the animal and human evidence. One particular aspect of communication needs to be emphasized here. Many people are interested in whether water fluoridation to prevent tooth decay poses a threat to human neurodevelopment and cognition. Although the monograph provides some discussion of dose–response relationships, NTP did not conduct a formal dose–response assessment and needs to state clearly that the monograph is not designed to be informative regarding decisions about fluoride concentrations for water fluoridation.

ANIMAL EVIDENCE

The monograph presents a systematic review of animal studies of fluoride exposure related to learning and memory that were published from 2015 to August

20, 2019, as an update to the NTP systematic review published in 2016. Examination of the animal studies published since 2015 led NTP to conclude that the animal data are inadequate to support conclusions on human cognitive effects.

The committee has serious concerns about the risk-of-bias evaluations of the animal literature and whether they identified important threats to internal validity that are specific to neurobehavioral outcomes in animal tests. The guidance in the protocol touched on some of the threats, but insufficient details appear to have been provided to ensure a rigorous, consistent evaluation of neurobehavioral studies. Specifically, the committee had concerns about the risk-of-bias evaluations for attrition, outcome assessment, and statistical analyses. It also found one element—maternal, fetal, and pup toxicity—that did not appear to have been adequately captured in the risk-of-bias criteria. Although severe postnatal toxicity was mentioned in risk-of-bias evaluations of some studies, it is unclear whether maternal, fetal, and pup toxicity was routinely assessed for all studies. Such effects can seriously confound interpretation of neurodevelopmental effects. Overall, the committee found that some studies cited in the monograph had severe methodologic shortcomings that could warrant exclusion from the body of evidence.

NTP justifies its conclusion that the animal evidence is inadequate on the grounds that it is not possible to separate cognitive effects from effects on locomotor activity. Although locomotor activity can affect learning and memory outcomes, it has been demonstrated many times that the presumed influence of locomotor activity on learning and memory does not occur. Thus, the committee does not agree with NTP's rationale for dismissing the animal evidence and finds that it is a mistake to dismiss studies of learning and memory because of minor, brief locomotor-activity changes or when other assessments can rule out confounding locomotor effects in cognitive assessments.

Given the serious concerns raised by the committee in the present report, NTP will need to decide whether it should reanalyze the animal evidence. The committee cautions, however, that given the poor quality of the animal studies that it reviewed, revising the systematic review to address the concerns highlighted might not affect the ultimate finding that the animal evidence is inadequate to inform conclusions about fluoride exposure and neurodevelopmental and cognitive effects in humans.

HUMAN EVIDENCE

NTP based its conclusion in the monograph primarily on human evidence. NTP considered the human evidence to be "relatively robust" and evaluated the association of fluoride exposure and neurodevelopmental and cognitive effects as reported in 82 publications. Although it evaluated all publications, its confidence in its conclusion is primarily based on the studies that were rated as having a lower risk of bias; NTP concluded that studies rated with a higher risk of bias did not

affect its confidence in its hazard conclusion. The committee had substantive concerns regarding NTP's evaluation of the human evidence as noted below.

The strategy used for the literature search indicated that NTP used FAN as a source to identify relevant literature. The process by which FAN identified and selected studies is unclear, and that uncertainty raises the question of whether the process could have led to a biased selection of studies. Such a concern raises the need for a formal evaluation of any potential bias that might have been introduced into the literature-search process. Another issue with the literature is that it appears that multiple publications are based on a single study and thus should not be considered independent studies. That lack of independence needs to be addressed in evaluating the findings and conclusions.

Several issues in the evaluation of risk of bias of individual studies were identified. First, there appeared to be inconsistent application of the risk-of-bias criteria across studies, perhaps stemming from differences in the approaches presented in the protocol and monograph. Second, the committee identified many cases in which NTP's evaluation of confounding was insufficient, difficult to understand, or applied inconsistently across studies. NTP should develop clear criteria that are defined in the protocol to identify critical confounders and, if these are not consistently applied to individual studies, explain why some potential confounders are considered to be of greater importance in some studies and not others. NTP should also address critical aspects of confounding, such as magnitude and directionality. Third, NTP noted the possibility of exposure misclassification in several cases but did not discuss its likely magnitude and direction and did not discuss it in the context of whether a given study reported an association. The failure to address exposure misclassification thoroughly and consistently raises the question of whether NTP's evaluations were sufficient and supported its conclusion. Fourth, it is imperative to protect examiners from information about exposure that could bias their administration and interpretation of outcome assessments, especially when they are assessing cognition or other neurobehavioral outcomes in human studies. Several studies reviewed by NTP did include information on techniques of blinding of examiners, but many did not. Because failure to blind examiners might result in a high risk of bias of study results and conclusions, NTP should consider this aspect more carefully when assessing the risk of bias of human studies. Fifth, NTP in some cases classified studies as having a low risk of bias when the measure of the neurodevelopmental and cognitive outcome was seriously flawed. Given the importance of that outcome in determining whether fluoride is hazardous, its proper measurement should be considered more carefully. Finally, the committee is concerned that the studies included in the systematic review did not undergo rigorous statistical review. That flaw is problematic because some of the studies identified as having low risk of bias did not adequately account for the hierarchical structure of their data or had errors in their summary statistics—faults that compromised their internal validity.

The committee also identified several issues with the analysis, summary, and presentation of the data. A key conclusion of the monograph is that the results of the epidemiologic studies consistently show a positive association. Although

the desire to provide a simple summary of a complex array of evidence is understandable, doing so requires comparing studies that have similar parameters, and this was not done in the monograph. In fact, the studies that are reviewed in the monograph used various measures of fluoride exposure and analytic techniques and evaluated neurodevelopmental and cognitive outcomes at different developmental times. The committee recognizes that drawing conclusions always requires aggregating or summarizing data that have some degree of heterogeneity among other considerations, but the monograph should juxtapose results across broadly comparable studies and use that information to provide a text summary of the patterns observed. If comparing "like to like" results yields consistent results across all measures, ages, exposure sources, statistical approaches, and exposure ranges, taking random error into account, that will indeed warrant a statement that results consistently show adverse effects. However, the monograph does not provide the evidence in a manner that leads to that conclusion. The committee notes that NTP did not conduct a meta-analysis. Given that meta-analysis is a useful tool for aggregating and summarizing data and analyzing comparable studies, the committee strongly recommends that NTP reconsider its decision not to perform one.

Lastly, the discussion section of the monograph provides an informal assessment of the evidence with regard to exposure and concludes that adverse health effects are observed largely in association with exposures above those associated with water fluoridation. The basis of that conclusion is not apparent and seems to contradict the earlier assertion that nearly all the studies are positive, including ones that evaluated groups exposed to lower concentrations. More important, as noted above, this discussion gives a false impression that NTP conducted a formal dose–response assessment. NTP should be clear that the monograph cannot be used to assess what concentrations of fluoride are safe.

NTP CONCLUSION

The monograph "concludes that fluoride is presumed to be a cognitive neurodevelopmental hazard to humans. This conclusion is based on a consistent pattern of findings in human studies across several different populations showing that higher fluoride exposure is associated with decreased IQ or other cognitive impairments in children." The committee was tasked with assessing whether NTP satisfactorily supports its conclusion. Given the issues raised by the committee regarding the analysis of various aspects of some studies and the analysis, summary, and presentation of the data in the monograph, the committee does not find that NTP has adequately supported its conclusion. That finding does not mean that the conclusion is incorrect; rather, further analysis or reanalysis as noted in the present report is needed to support conclusions in the monograph.

1

Introduction

Water fluoridation has long been hailed as an effective method of reducing dental caries (tooth decay). Over the years, however, people have raised concerns about adverse health effects of fluoride exposure. Of particular concern are results of epidemiologic studies—typically conducted in regions that have high naturally occurring fluoride—that have reported neurodevelopmental and cognitive effects in humans. That concern and a nomination from the Fluoride Action Network (FAN) prompted the National Toxicology Program (NTP) Office of Health Assessment and Translation to undertake a systematic review to evaluate the evidence of adverse neurodevelopmental and cognitive effects of fluoride exposure in humans. To ensure the integrity of its evaluation, NTP asked the National Academies of Sciences, Engineering, and Medicine (the National Academies) to review its monograph *Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects* (NTP 2019).[1] As a result of that request, the National Academies convened the present committee, which prepared this report.

FLUORIDE TOXICITY

Water fluoridation in the United States began in 1945 as a public-health practice to prevent dental caries. In 1962, the US Public Health Service recommended optimal fluoride concentrations of 0.7–1.2 mg/L; it revised its recommendation to 0.7 mg/L in 2015 (Gooch 2015). State and local governments, however, ultimately decide whether to fluoridate water systems. From the outset, the practice of fluoridating water systems has been controversial (NRC 2006), primarily because of the adverse health effects that have been associated with fluoride exposure over the years.

In 2006, the National Academies released the report *Fluoride in Drinking Water: A Scientific Review of EPA's Standards* (NRC 2006), which reviewed the scientific literature on fluoride exposure and human health effects (see Box 1-1). That report found that chronic exposure to fluoride is associated with enamel fluorosis and with bone weakening that could increase the risk of fractures. However, the evidence on several outcomes was not sufficient for the committee to reach conclusions; neurotoxicity was one such outcome. A few epidemiologic studies

[1]Hereafter referred to as the monograph.

BOX 1-1 Fluoride in Drinking Water: A Scientific Review of EPA's Standards

NRC (2006) reviewed the US Environmental Protection Agency (EPA) drinking-water standards for fluoride as a natural contaminant of the public water supply, not as an artificial additive to water supplies for dental-health protection. EPA drinking-water standards are the maximum-contaminant-level goal (MCLG) and the maximum contaminant level (MCL). The MCLG is a health goal that is set at a concentration at which no known or expected adverse health effects are expected to occur within adequate margins of safety. The enforceable drinking-water standard is the MCL, which is set as close to the MCLG as possible after consideration of such factors as available treatment technology and costs. The MCLG and the MCL for fluoride are both 4 mg/L. The committee that wrote the report was unanimous in its conclusion that the MCLG of 4 mg/L should be lowered because it puts children at risk for severe enamel fluorosis. The majority of the committee also concluded that exposure to fluoride at the MCLG is likely to pose a risk of bone fractures.

indicated IQ deficits in children exposed to fluoride at 2.5–4 mg/L in drinking water. However, the committee that prepared the 2006 report concluded that "the studies lacked sufficient detail…to fully assess their quality and relevance to the U.S. populations, [but] the consistency of the results appears significant enough to warrant additional research on the effects of fluoride on intelligence" (NRC 2006, p. 8).

THE NTP FLUORIDE MONOGRAPH

Since the National Academies report (NRC 2006) was released, additional scientific research has been conducted on the association between fluoride exposure and neurodevelopmental and cognitive health effects. In 2016, NTP published the results of a systematic review that examined the effects of fluoride exposure on learning and memory in animals (NTP 2016). NTP found "low to moderate level-of-evidence that suggests adverse effects on learning and memory in animal[s] exposed to fluoride" at concentrations higher than 0.7 ppm (NTP 2016, p. vii). NTP noted that few studies that examined effects near concentrations of 0.7 ppm were available, that confidence in the results of available studies was reduced because of confounding and risk-of-bias issues, and that further research was needed.

Over the last decade, epidemiologic studies of the effects of fluoride exposure on neurodevelopment and cognition have also been conducted. Given those studies and a nomination from FAN, NTP conducted a systematic review of the

evidence on fluoride exposure and neurodevelopmental and cognitive health effects and released its monograph in 2019 (NTP 2019).[2] Although a primary focus was on the human evidence, the systematic review evaluated animal studies that had been published since the 2016 NTP report and mechanistic studies that might be able to shed light on a possible pathway for fluoride exposure to cause neurodevelopmental or cognitive health effects. The monograph concluded that "fluoride is presumed to be a cognitive neurodevelopmental hazard to humans. This conclusion is based on a consistent pattern of findings in human studies across several different populations showing that higher fluoride exposure is associated with decreased IQ or other cognitive impairments in children" (NTP 2019, p. 2). Although NTP did not conduct a formal dose–response assessment, it noted that effects on cognitive neurodevelopment were inconsistent at concentrations of about 0.03–1.5 ppm. NTP (2019) also stated that the evidence of cognitive effects in adults was inadequate, that the animal evidence was inadequate to support conclusions about cognitive effects, and that the possible mechanisms for the noted effects "are not well characterized." Given the importance of the findings, NTP asked the National Academies to review its monograph.

STATEMENT OF TASK

The committee that was convened as a result of the NTP request included experts in toxicology, epidemiology, neurodevelopment, systematic review, and statistics. Appendix A provides biographic information on the committee. The committee was asked to review the monograph and ultimately to assess whether NTP's conclusions are supported by the evidence provided in it. The verbatim statement of task is provided in Box 1-2.

THE COMMITTEE'S APPROACH TO ITS TASK

The committee held several teleconferences and one in-person meeting, which included an open session at which the committee heard from the sponsor and interested stakeholders. As part of its evaluation, the committee reviewed key scientific studies from the monograph and considered materials submitted to the committee by interested parties. It is important to note that the committee did not conduct its own independent evaluation of the evidence, and it did not conduct a data audit (that is, review all the data reported in the monograph to ensure that it had been reported correctly), although it did review some key literature to enable

[2]It is important to note that NTP monographs evaluate the evidence that a given exposure causes adverse health effects. As noted by NTP, the monographs might provide hazard conclusions depending on the assessment goals and available evidence (see https://ntp.niehs.nih.gov/whatwestudy/assessments/index.html). They typically do not include formal dose–response assessments, and they are not risk assessments or risk–benefit assessments. Therefore, they should not be used to reach conclusions on appropriate exposure guidance levels or standards.

BOX 1-2 Statement of Task

An ad hoc committee of the National Academies of Sciences, Engineering, and Medicine will review the *National Toxicology Program (NTP) Monograph on Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects*. The committee will provide an overall critique of the draft monograph and address the following questions:

- Has the systematic review protocol been followed and modifications appropriately documented and justified?
- Does the monograph accurately reflect the scientific literature?
- Are the findings documented in a consistent, transparent, and credible way?
- Are the report's key messages and graphics clear and appropriate? Specifically, do they reflect supporting evidence and communicate effectively?
- Are the data and analyses handled in a competent manner? Are statistical methods applied appropriately?
- What other significant improvements, if any, might be made in the document?
- Does the scientific evidence in the NTP monograph support NTP's hazard category conclusions for fluoride in children and adults?

its review. The committee evaluated whether presentation of the evidence in the monograph supported NTP's conclusions and focused primarily on the human and animal evidence.

ORGANIZATION OF THIS REPORT

The report is organized into four chapters and one appendix. Chapter 2 provides the committee's review of the methods and overall presentation. Chapters 3 and 4 provide the committee's evaluation of NTP's presentation and assessment of the animal and human evidence, respectively. Appendix A provides the biographic information on the committee.

REFERENCES

Gooch, B.E. 2015. U.S. Public Health Service Recommendation for Fluoride Concentration in Drinking Water for the Prevention of Dental Caries. Public Health Reports 130: 318-331.

NRC (National Research Council). 2006. Fluoride in Drinking Water: A Scientific Review of EPA's Standard. Washington, DC: The National Academies Press.

NTP (National Toxicology Program). 2016. NTP Research Report: Systematic Review of the Effects of Fluoride on Learning and Memory in Animal Studies. Research Triangle Park, NC: National Toxicology Program. Research Report 1.

NTP. 2019. Draft NTP Monograph on the Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects. Office of Health Assessment and Translation, Division of the NTP, National Institute of Environmental Health Sciences, National Institutes of Health, US Department of Health and Human Services.

2

Methods and Presentation

Starting in 2011, the National Toxicology Program (NTP) Office of Health Assessment and Translation (OHAT) began developing and incorporating systematic review methods into literature evaluations to assess scientific evidence of human health effects of exposures to environmental chemicals, physical substances, or mixtures (Birnbaum et al. 2013; Rooney et al. 2014). That effort was part of a cultural change within the environmental-health field in which approaches to evaluation of scientific evidence resulted predominantly in expert-based narrative reviews. However, mounting empirical evidence that narrative reviews generally lacked the ability to evaluate evidence in a rigorous, systematic, transparent, and reproducible manner indicated that more rigorous approaches to evidence-based decision-making were needed (Reenie and Chalmers 2009; NRC 2011; Woodruff and Sutton 2011). As a result, several agencies and institutions have undertaken the development and implementation of systematic review methods to address environmental questions (EFSA 2010; Woodruff and Sutton 2011, 2014; Murray and Thayer 2014).

NTP OHAT systematic review methods are described in several documents. First, the OHAT handbook on systematic review (published in 2015 and updated in 2019) represents the "standard operating procedures" for how systematic review and evidence integration are to be conducted for OHAT literature-based assessments (NTP 2015, 2019a,b).[1] Those operational guidelines are based largely on empirically tested approaches and expert input from various fields, such as the clinical sciences, including Cochrane (Higgins et al. 2019), Grading of Recommendations Assessment, Development, and Evaluation (Guyatt et al. 2008), the Navigation Guide (Woodruff and Sutton 2011, 2014), the Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental

[1]Although the protocol for the fluoride monograph refers to both versions of the OHAT handbook, the committee assumes that it was based on the updated version given that it incorporates a revised figure (NTP 2017, Figure 3, p. 20) that was provided in the updated version.

Studies (CAMARADES),[2] and the Systematic Review Centre for Laboratory Animal Experimentation (SYRCLE).[3]

For NTP OHAT literature-based assessments, such as the monograph evaluated in the present report, a protocol is developed and shared publicly before beginning the assessment. A protocol is the detailed plan that is to be followed in a specific systematic review and describes the rationale, the objectives of the review, and the conduct of each step of the review (NTP 2015; Higgins et al. 2019). The steps include problem formulation, which results in the development of the Population, Exposure, Comparator, and Outcomes (PECO) statement; development of inclusion and exclusion criteria for study selection; a search of the literature; extraction of data from included studies; critical appraisal of studies for risk of bias; synthesis of results from included studies; and hazard identification by integration of the evidence from human and animal studies and consideration of supporting evidence from mechanistic studies. Ideally, the protocol should follow guidance provided by the OHAT handbook but include details specific to the given systematic review, such as how to rate risk of bias in assessing outcomes of neurodevelopmental and cognitive effects as in the case here.

The methods section in the monograph should also describe how a specific systematic review was conducted. Ideally, the methods described in the monograph should align with the details outlined in the protocol or should transparently and explicitly document, describe, and justify any deviations from the protocol. In the present chapter, the committee provides its assessment of the methods and overall presentation.

GENERAL ASSESSMENT

The committee noted several strengths of the monograph. As noted, the protocol is a critical component of a systematic review and ideally minimizes reviewer bias, allows feedback at early stages of the systematic review, and transparently highlights any changes made as the systematic review process unfolds (IOM 2011). The protocol for the monograph contains descriptions of each step of the systematic review and clearly outlines several protocol revisions, including the date and justification of each change. Furthermore, changes in the protocol are clearly indicated so that both the original text and the modified text are readily

[2]CAMARADES provides a supporting framework for groups involved in the systematic review and meta-analysis of data from experimental animal studies. See http://www.dcn.ed.ac.uk/camarades/.

[3]SYRCLE focuses on the execution of systematic reviews of animal studies aimed at more evidence-based translational medicine. See https://www.radboudumc.nl/en/research/departments/health-evidence/systematic-review-center-for-laboratory-animal-experimentation.

apparent. NTP appears to be adhering to best practices for systematic reviews with respect to the availability and documentation of such a protocol before initiation of a review.

The committee also was impressed by the availability of systematic review data in the interactive, freely available program Health Assessment Workspace Collaborative (HAWC).[4] Given the plethora of data extracted from studies and the risk-of-bias ratings and justifications, interactive programs increase the ease with which an independent reviewer can explore the data in more detail without being limited to the graphics and tables provided in the report itself. The committee appreciated that most of the tables and figures in the monograph were available with additional study details and interactive graphics in HAWC.

The committee, however, had some overarching concerns regarding the protocol, data presentation, and communication that are described in the following sections, and it provides some suggestions for improvements.

PROTOCOL

The protocol for the systematic review described in the monograph was published on NTP's Web site in June 2017 and made available for public comment. It was reviewed several times by technical advisers selected for their expertise on this topic. In general, it describes the overall systematic-review process and clearly outlines modifications that were made during the review. The committee, however, identified several issues associated with the protocol. First, the role of the OHAT handbook in developing the protocol is unclear. Second, important details are missing from the protocol. Third, inconsistencies between the protocol and the monograph raise concerns. Those issues are discussed further below; other issues associated with execution of the protocol are discussed in Chapters 3 and 4. Given the issues raised here and in later chapters, the committee finds that there are some deficiencies in the protocol and its execution.

Role of the OHAT Handbook

As discussed, the OHAT handbook outlines "standard operating procedures for systematic review and evidence integration for conducting OHAT literature-based assessments" (NTP 2019a, p. v) and is intended as a "living document" that is continually updated to reflect refinement and modifications of the OHAT approach. However, the protocol scarcely refers to the OHAT handbook and does not discuss its role in developing the protocol. Specifically, the only references to the OHAT handbook in the protocol are noted below.

- A statement that "the systematic review will be based on guidance outlined in the Office of Health Assessment and Translation (OHAT) Handbook for Conducting a Literature-Based Assessment" (NTP 2017, p. 3).

[4]See https://hawcproject.org/user/login/?next=/portal/.

- Supporting descriptions of "tier 3" studies that have overall very serious risk-of-bias concerns for individual epidemiology studies (NTP 2017, pp. 8, 10).
- Further guidance for assessing confidence in the body of overall evidence (NTP 2017, p. 13).

Thus, the role of the OHAT handbook in developing the protocol is unclear. That ambiguity leads to concerns about the lack of detail in the protocol and about apparent conflicts between the methodologic approach in the protocol and the OHAT handbook itself. Several examples of that ambiguity and the associated concerns are provided below.

- *Nomination history*. According to the OHAT handbook, this section should describe "the history of the nomination…steps the NTP has taken to solicit feedback on the topic under consideration, including *Federal Register* notices, requests for information in the *NIH Guide for Grants and Contracts*, outreach to federal agencies on the NTP Executive Committee, or outreach to other divisions within NIEHS [National Institute of Environmental Health Sciences]" (NTP 2019a, p. 12). This section should also provide a summary of any comments received during the comment periods. However, the protocol mentions only briefly when the topic was nominated and when it was presented to the NTP Board of Scientific Counselors (NTP 2017, p. 5). Some additional detail is provided in the section "Nominations to NTP" (NTP 2017, p. 2) and in the monograph itself, which acknowledges that the nomination was from the Fluoride Action Network, but the protocol does not address the extent of topics as outlined in the OHAT handbook.
- *Problem formulation*. According to the OHAT handbook, this section should "describe and document major decisions made during scoping and problem formulation. It should also describe how key scientific issues will be addressed in the evaluation. Problem formulation activities include discussions of the evaluation design team, preparation of scoping reports and any external activities, such as concept review by the NTP Board of Scientific Counselors, public comment, or webinars, listening sessions, or workshops undertaken to solicit input on specific scientific or technical issues" (NTP 2019a, p. 16). However, that information appears to be missing from the protocol. In the monograph, the section "Problem Formulation and Protocol Development" contains a list of problem-formulation steps, including input from the NTP Board of Scientific Counselors and a review of the draft protocol by technical advisers (NTP 2019c, p. 5). Notably missing is a discussion of opportunities for public engagement and comment except for acknowledgment that the protocol has been publicly available on NTP's Web site since June 2017. The OHAT Web site appears to indicate that there were several public-comment periods in 2015 and 2016, but they are not reported or discussed in the protocol.

- *Development of PECO statement*. The OHAT handbook includes a section "Key Questions and Analytical Framework" that guides development of the PECO statement, but a similar section is not included in the protocol or the monograph.
- *Screening and data extraction*. In the protocol, four NIEHS staff and numerous ICF contractors are identified as involved in the screening step. The OHAT handbook states that if a contractor is used, a second reviewer should be an NTP staff member. It is unclear whether that guidance was followed. For data extraction, the protocol does not mention training for data extractors or pilot testing of all team members as recommended in the OHAT handbook.

To increase transparency, NTP should clearly describe the role of the OHAT handbook in developing the systematic-review protocol primarily to set expectations for how closely the process described in the handbook will be followed in the protocol and eventually the systematic review. That would help to address concerns about information that appears to be missing from the protocol or about conflicts between the protocol and the handbook.

Important Details

The presentation by NTP OHAT staff to the committee on November 6, 2019, indicated that the protocol for the monograph is intended to serve as a standalone document—that is, the protocol should contain all details relevant to the conduct of the systematic review. That position might be due partly to the fact that the OHAT handbook will change, as it states: "the procedures are a living document with the expectation that approaches will be updated as methodological practices are refined and strategies identified that improve the ease and efficiency of conducting a systematic review" (NTP 2019a, p. v). The OHAT handbook has already undergone one recent revision (NTP 2019b) and will likely undergo several more. Thus, it might be best if each systematic-review protocol could stand alone as an independent document that contains all the information necessary for understanding of the planning and conduct of the review. The committee, however, acknowledges that it would also be satisfactory to cite the appropriate OHAT handbook versions in the protocol for sections in which the details regarding the process align with the handbook to limit the need to repeat information from the handbook in the protocol.

Any details pertaining to the conduct of a review should be planned beforehand and described in the protocol. The committee found that many important details were missing from the protocol, although some of the information was contained in the OHAT handbook or the monograph. The committee recommends that the details be included in the protocol for transparency. Examples of important details missing in the protocol are as follows:

- *Updating animal literature search.* One of the specific aims of the systematic review was to update the experimental animal literature cited in NTP's systematic review of the animal evidence on effects of fluoride on learning and memory (NTP 2016). However, details on the procedures used to update the literature search were lacking. The protocol provides database search strategies in Appendix 1, but these are not specific to the evidence stream and list only an end date for the search (December 19, 2016, for PubMed). The search strings also appear to differ from those used in the previous animal systematic review (NTP 2016). Thus, it appears that the literature search for animal studies is not a "re-execution"[5] of the original search but rather an "update"[6] of the search. However, it is unclear how the updated search specifically for animal studies was conducted and whether the modifications in search strategy resulted in the identification of new studies published before 2016. The monograph discusses the search strategies to some extent by stating that "literature searches for this systematic review were conducted independent of the literature search conducted for the NTP (2016) report using a similar strategy. As relevant animal studies published prior to 2015 were identified in the NTP (2016) assessment, the focus of the literature searches for this systematic review was to identify relevant animal studies that were published since completion of the literature searches for the NTP (2016) assessment" (NTP 2019c, p. 8). The specific procedure for updating the literature search for animal studies should be transparently outlined in the protocol with sufficient details to allow independent reproduction of the search.
- *Evidence selection criteria*. The protocol includes a detailed PECO statement but does not include explicit inclusion and exclusion criteria. That omission is critical because, although a PECO statement forms the basis of the criteria, detailed criteria offer greater clarity for understanding and documenting the screening process for identifying relevant studies. The criteria also increase transparency for potential reproduction of the review or facilitate updates to incorporate newer information. The example provided in the OHAT handbook illustrates how detailed inclusion and exclusion criteria provide greater clarity than simply a PECO statement alone (NTP 2019a. pp. 13, 17).
- *Screening for inclusion*. Studies were screened for inclusion by using a structured form in SWIFT-Active Screener, a machine-learning software program used to rank studies for screening. The National Academies has stated that automated screening procedures can facilitate efficiencies in the process and that incorporation of software tools, such as SWIFT-Active Screener, can help to achieve that goal (NRC 2014; NASEM 2018). However, those tools are relatively new and have not undergone rigorous evaluation or validation. Specifically, to the committee's knowledge, they have not been validated for screening studies for inclusion in systematic reviews. Furthermore, screening up to 98% inclusion means that as many as 2% of the 13,023 studies excluded on the basis of the

[5]See https://handbook-5-1.cochrane.org/chapter_3/3_4_2_1_re_executing_the_search.htm.
[6]See https://handbook-5-1.cochrane.org/chapter_6/6_4_12_updating_searches.htm.

SWIFT algorithm in this systematic review—260 studies—could be relevant according to title and abstract screening but missed in the initial screening. Given the large number of studies screened for this systematic review, that is not an insignificant number, although the committee notes that not all the studies would likely be deemed relevant in the full-text screening step. The OHAT handbook mentions the SWIFT text-mining and machine-learning tools but does not justify or cite why 98% estimated recall is considered sufficient. The committee recommends that the protocol discuss the basis of that decision and potentially conduct a sensitivity analysis to determine the effect of that cutoff on the overall findings (for example, by reviewing a random subset of the studies excluded on the basis of the SWIFT algorithm to identify the number of potentially missed references).

- *Reporting excluded studies*. It was decided in a May 2019 protocol revision not to list excluded studies because the number of studies excluded by using SWIFT Active during title and abstract screening was large. The OHAT handbook indicates that the list of included and excluded studies should be posted on the project's Web site when screening has been completed to provide an opportunity for public review of the literature considered for evaluation. There is no mention of a size-cutoff criterion. The committee finds that further justification of the decision is warranted; this pertains to a list of 9,667 references screened and excluded and 13,023 references not screened—not an unreasonable number to present. In particular, the list of 2% unscreened studies based on 98% recall in SWIFT Active would be particularly appropriate to include because there is a chance that up to 260 studies could be missed as noted above.
- *Screening and data extraction*. The protocol lacks details relevant to screening and data extraction. For example, the protocol (NTP 2017, p. 29) describes the ideal evaluation team by stating that the "team members should have at least a master's degree or equivalent level of experience in epidemiology, toxicology, environmental health sciences, or a related field." However, it is unclear whether that criterion was met by the members of the review team inasmuch as only their names and affiliations are provided. Furthermore, more ICF contractors are listed in the monograph than in the protocol, but it is unclear when and why they were added. The committee recommends that the expertise and experience of all team members be provided for transparency to ensure that the review team has been established with expertise and experience appropriate for conducting the systematic review and recommends that OHAT guidance regarding the screening and data extraction process be followed.
- *Data synthesis.* The protocol does not include details about the planned conduct of statistical analyses, for example, models for meta-analyses, meta-regression, sensitivity analyses, or statistical evaluation of publication bias. The protocol contains only a section that discusses consideration for pursuing a narrative or quantitative evidence synthesis and addresses heterogeneity in the available evidence. One assumes that the approach outlined in the OHAT handbook was applied, but that is not explicitly stated. Given that data analyses can vary with

specific study questions, it would be appropriate to include a section on data analysis in the protocol. That approach is consistent with Cochrane and other protocols for environmental health.

- *Rating evidence*. The protocol contains details for each factor that contributes to increasing or decreasing confidence in the body of evidence (see Table 4 "Key Factors when Considering Whether to Downgrade or Upgrade across a Body of Evidence," NTP 2017, p. 14ff), but the committee did not find that these descriptions were sufficient to ensure reproducibility or transparency of the process. For example, the downgrade factor of "risk of bias" includes only a list of the critical factors that potentially contribute to high overall risk-of-bias ratings. However, that guidance does not sufficiently outline the criteria that make it appropriate to downgrade for risk-of-bias concerns. NTP should clearly define each factor, including key considerations that warrant upgrading or downgrading the body of evidence. If the factors for upgrading and downgrading the body of evidence align with criteria provided in the OHAT handbook, that should be explicitly stated, and the appropriate version of the handbook should be cited to indicate where additional details might be found.

Consistency

Some details outlined in the protocol appeared inconsistent with methods ultimately implemented in the monograph. Given the critical role of the protocol in the design and implementation of a systematic review, those potential discrepancies were concerning. Changes in the protocol are common, as illustrated by several transparently documented modifications of the protocol, but inconsistencies between the protocol and the monograph raise concerns regarding the rationale for specific changes and why they were not documented in the protocol.

- *Consideration of mechanistic data.* The protocol states that if "mechanistic data fail to provide support for biological plausibility of the relationship between exposure and the health effect, the hazard identification conclusion may be downgraded…from that initially derived" (NTP 2017, p. 22). However, the monograph differs by stating that "if mechanistic data provide strong opposition for biological plausibility of the relationship between exposure and the health effect, the hazard identification conclusion may be downgraded…from that initially derived" (NTP 2019c, p. 16). The committee finds that the latter approach is more appropriate because failing to provide support of a relationship is fundamentally different from providing support against a relationship. The committee concludes that there must be sufficient mechanistic evidence to warrant downgrading a hazard conclusion, not simply lack of evidence that supports a relationship.
- *Outcome assessment*. The protocol does not specify any subgroup or sensitivity analyses to be conducted. In the monograph, however, the decision to evaluate child and adult outcomes separately is presented. The protocol did not explicitly discuss or justify that decision or provide a definition to guide analysis.

The monograph does discuss how the NTP (2016) report considered the two age groups for experimental animals separately and indicates that this systematic review has mirrored such an approach (NTP 2019c, p. 4). However, the protocol should indicate when the decision was made, provide justification for it, and discuss specifics of the approach, for example, defining the age ranges that constitute "adult" and "child." The protocol also should specify how evidence from studies that evaluate mixed populations (those containing children and adults) were incorporated. In general, any planned subgroup or sensitivity analyses should be described and justified in the protocol.

- *Confounders adjustment*. In the protocol, the two critical confounders were identified as the potential for co-exposures (arsenic and lead) and iodine sufficiency. In the monograph, however, iodine deficiency or excess is listed as a potential confounding variable that might be considered important but not necessary; this is a major difference from the protocol, and it is unclear why it changed. Figure 6 in the monograph identifies key confounders as age, sex, arsenic, and socioeconomic status. Again, that is a discrepancy from what is discussed earlier in the monograph and in the protocol. Furthermore, a checkmark in Figure 6 indicates that the factor was considered and might have been adjusted for in the final model, but the criterion in the protocol for a low rating is that the study must provide quantitative summaries of the covariate *and* adjust for it in the analysis. Such discrepancies should be reconciled.

DATA PRESENTATION

As stated in Chapter 1, the committee did not conduct a data audit, but it did find some minor issues with data presented in the monograph. Some of the issues appear to be minor errors; for example, the number of references is incorrectly reported in Figure 4 "Study Selection Diagram" (NTP 2019c, p. 18). During the title and abstract screening, 9,667 references were screened and excluded; 13,023 references were not screened, on the basis of the SWIFT algorithm; and 807 references were included for full text review. Those numbers sum to 23,497—30 more studies than reported as screened in the figure (23,467).

Other presentation issues are related to enhancing the utility of the data presentation. For example, most tables and figures in the monograph are organized alphabetically by study author last name. That approach does not convey the information in a meaningful format. It might be better to organize the studies in a more informative way; for example, risk-of-bias tables could be organized by risk-of-bias ratings (that is, studies that have the most green "ratings" first and studies that have fewer such ratings thereafter) or by stratifying the studies according to critical risk-of-bias domains similar to the example provided in the OHAT handbook (NTP 2015, p. 39). Because critical domains[7] might cause a

[7]Key domains for humans include confounding, exposure characterization, and outcome assessment, and key domains for animals include randomization, exposure characterization, outcome assessment, and litter effects for developmental studies.

study to be excluded from the analysis, highlighting them in each table would help readers to interpret the overall risk-of-bias ratings. As another example, Table 6, "Studies on Neurodevelopmental and Cognitive Function in Humans," also arranged studies alphabetically by study author last name (NTP 2019c, p. 22ff). An alternative suggestion would be to categorize studies by age cutoffs because the outcomes are evaluated separately for children and adults. Furthermore, NTP should include the detection limit of exposure measures for each study in Table 6. Organization of data is discussed further in Chapters 3 and 4.

COMMUNICATION

The committee noted that the monograph might benefit from improved communication in several respects. First, many people are interested in the question of whether water fluoridation to prevent dental decay poses a threat to cognition and neurodevelopment. Although the monograph includes some discussion of dose–response relationships, NTP did not conduct a formal dose–response assessment that could inform a discussion on water fluoridation. NTP needs to state clearly that the monograph is not designed to be informative with respect to decisions about the concentrations of fluoride that are used for water fluoridation. That point should be reiterated at the end of the monograph with some indication that its evaluation of the literature is focused on hazard identification of fluoride and that it does not draw any conclusions regarding drinking-water fluoridation or other fluoride sources, such as toothpaste or other dental treatments. Although NTP does not explicitly claim that it has done something other than hazard identification, the context into which the monograph falls calls for much more carefully developed and articulated communication on this issue.

Second, the monograph lacks details on the process of evaluating confidence in the body of evidence. As discussed earlier, the protocol lacks sufficient definitions for level-of-evidence descriptors as they pertain to the specific study question addressed in the monograph. However, the ratings applied for each confidence factor (Table 7, NTP 2019c, p. 51) also lack any justification or discussion. For example, NTP's discussion of the overall risk-of-bias rating (NTP 2019c, p. 28) states that the confidence rating was not downgraded because 20 studies had little or no risk-of-bias concerns. However, NTP also states that the remaining human studies had probably high or definitely high risk of bias for at least two key considerations (exposure characterization, outcome assessment, or confounding factors). Thus, the rationale for not downgrading risk of bias is not entirely clear. NTP should consider supplementing Table 7 with clear justification for each confidence factor rationale for and why upgrade and downgrade factors were not applied for any of the human evidence.

Finally, there is little discussion of the process for obtaining missing or additional information from study authors. The monograph states that authors of included studies "were queried by email to obtain missing information and responses received were used to update risk-of-bias ratings" (NTP 2019c, p. 12). The author responses and changes in risk-of-bias ratings were documented in the

HAWC program. Obtaining additional information from study authors is critical for systematic review because it can minimize the effect of reporting bias vs other aspects of bias related to the design and implementation of studies. However, the monograph lacks clear documentation of how many authors were contacted, which authors were contacted, how many responded, and how risk-of-bias ratings were updated generally. Although that is tracked in HAWC with the actual risk-of-bias ratings, it could be helpful to include overall summary statistics of this critical process for greater transparency in the monograph. At a minimum, risk-of-bias ratings or extracted data that are updated on the basis of information obtained from the study authors should be clearly indicated for the specific risk-of-bias ratings or data in all relevant tables and figures.

REFERENCES

Birnbaum L.S., K.A. Thayer, J.R. Bucher, and M.S. Wolfe. 2013. Implementing systematic review at the National Toxicology Program: Status and next steps. Environmental Health Perspectives 124(4):a108-a9.

EFSA (European Food Safety Authority). 2010. Application of systematic review methodology to food and feed safety assessments to support decision making. EFSA Journal 8(6):1-1690.

Guyatt G.H., A.D. Oxman, G.E. Vist, R. Kunz, Y. Falck-Ytter, P. Alonso-Coello, H.J. Schünemann, and GRADE Working Group. 2008. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 336(7650): 924-926.

Higgins J.P.T., J. Thomas, T. Li, M.J. Page, V. Welch, M. Cumpston, J. Chandler, and L. Mellor. 2019. Cochrane Handbook for Systematic Reviews of Interventions. Chichester, UK: John Wiley & Sons.

IOM (Institute of Medicine). 2011. Finding what works in health care: standards for systematic reviews. Washington, DC: National Academies Press.

Murray H.E., and K.A. Thayer. 2014. Implementing systematic review in toxicological profiles: ATSDR and NIEHS/NTP collaboration. Journal of Environmental Health 76(8):34-35.

NASEM (National Academies of Sciences, Engineering, and Medicine). 2018. Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation. Washington, DC: National Academies Press.

NRC (National Research Council). 2011. Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde. Washington, DC: National Academies Press.

NRC. 2014. Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, DC: National Academies Press.

NTP (National Toxicology Program). 2015. New OHAT Handbook for Conducting Systematic Reviews. Available: https://ntp.niehs.nih.gov/update/2015/1/ohat-handbook/index.html.

NTP. 2016. NTP Research Report: Systematic Review of the Effects of Fluoride on Learning and Memory in Animal Studies. Office of Health Assessment and Translation, Division of the NTP, National Institute of Environmental Health Sciences, National Institutes of Health, US Department of Health and Human Services.

NTP. 2017. Protocol for Systematic Review of Effects of Fluoride Exposure on Neurodevelopment.

NTP. 2019a. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Research Triangle Park, NC: National Institute of Environmental Health Sciences.

NTP. 2019b. Updates and Clarification to the OHAT Approach for Systematic Review and Evidence Integration Research Triangle Park, NC: National Institutes of Health. Available: https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookclarificationmarch2019_508.pdf.

NTP. 2019c. Draft NTP Monograph on the Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects. Office of Health Assessment and Translation, Division of the NTP, National Institute of Environmental Health Sciences, National Institutes of Health, US Department of Health and Human Services.

Rennie D., and I. Chalmers. 2009. Assessing authority. JAMA 301(17):1819-1821.

Rooney A.A., A.L. Boyles, M.S. Wolfe, J.R. Bucher, and K.A. Thayer. 2014. Systematic review and evidence integration for literature-based environmental health science assessments. Environmental Health Perspectives 122(7):711-718.

Woodruff T.J., and P. Sutton. 2011. Navigation Guide Work Group. An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. Health Aff (Millwood) 30(5):931-937.

Woodruff T.J. and P. Sutton. 2014. The Navigation Guide systematic review methodology: A rigorous and transparent method for translating environmental health science into better health outcomes. Environmental Health Perspectives 122(10):1007-1014.

3

Animal Evidence

The National Toxicology Program (NTP) monograph presents a systematic review of animal studies of fluoride exposure related to learning and memory that were published from 2015 to August 20, 2019, as an update of the agency's 2016 systematic review (NTP 2016). The 2016 review found a low-to-moderate level of evidence that learning and memory deficits occur in experimental animals exposed to fluoride, a finding that prompted NTP to conduct its own study (McPherson et al. 2018). On the basis of its updated systematic review, NTP changed its conclusion to a finding that the animal data are *inadequate* to inform conclusions on cognitive effects. This chapter reviews the major steps of the systematic review and how it was used to draw hazard-identification conclusions.

LITERATURE SEARCH AND SCREENING

In accordance with its protocol, NTP conducted a comprehensive literature search for animal studies of fluoride exposure and measures of learning and memory.[1] However, NTP's inclusion and exclusion criteria for screening the animal literature at the full text level are not documented in either the protocol or the monograph. As noted in Chapter 2, that omission is critical because detailed criteria offer greater clarity for understanding and documenting the process for identifying relevant studies.

RISK-OF-BIAS EVALUATION

A concern for the committee was whether NTP's risk-of-bias evaluations adequately captured important threats to internal validity that are specific to neurobehavioral outcomes in animal tests. The guidance in the protocol touched on

[1]Although the mechanistic evidence is considered separately from the animal evidence, and the committee focused on the animal evidence, the committee questioned whether the literature search strategy for mechanistic evidence was adequate to capture all relevant information, especially with respect to studies that analyzed data derived from new approach methodologies. For example, publications that report testing of large numbers of chemicals in which chemical names are present only in tables or supplemental files might not be captured.

some of the threats, but insufficient detail was provided to ensure a rigorous, consistent evaluation of neurobehavioral studies. Concerns regarding several risk-of-bias domains are discussed below.

Attrition

Findings of neurotoxicity or neurodevelopmental toxicity of any kind are seriously confounded at doses that result in excessive animal deaths. For example, Kinawy and Al-Eidan (2018)—a study considered by NTP—reported a 30% decrease in the number of offspring in a fluoride-treated group; this result raises serious concerns about the validity of the study's findings. Given its review of some key studies, the committee is concerned that high mortality was not adequately considered by NTP when it evaluated the animal studies for this risk-of-bias domain.

Confidence in the Outcome Assessment

The protocol lists examples of "well-established methods" for measuring particular outcomes and refers to meeting "standard protocols for each of these well-established methods" for a study to be rated as having a low risk of bias (NTP 2017, p. 65). The committee found, however, several examples of studies in which inappropriate testing procedures were followed or descriptions of methods were insufficient to evaluate their adequacy. That issue indicated to the committee that the risk-of-bias raters simply looked for a test name to determine whether the method was "reliable" without assessing whether methods were suitable to support confidence in the validity of the results. For example, some of the test periods in open-field tests of locomotor activity were too short (3–5 min) to provide reliable, reproducible, or meaningful results (see, for example, Balaji et al. 2015; Bartos et al. 2015; Pereira et al. 2011; Kivrak 2012; Sárközi et al. 2015; Zheng et al. 2016; Nageshwar et al. 2017; Nkpaa and Onyeso 2018; Sun et al. 2018; Wang et al. 2018). All those studies are described by raters as having used acceptable methods. Many published reports indicate that important chemical-related effects on motor activity are detected only after the first 3–5 min of testing (see, for example, Curran et al. 2011; Amos-Kroohs et al. 2015). For that reason, national and international guidance and guidelines for neurotoxicity and developmental-neurotoxicity testing recommend testing with automated systems that have activity sessions of at least 30 min (see, for example, EPA 1998a,b; OECD 2007; NAFTA 2016).

Another concern is that multiple studies provide incomplete descriptions of neurobehavioral test methods (see, for example, Zhu et al. 2017; Sharma et al. 2018; Raju et al. 2019). Behavioral testing methods, including such widely used tests as the Morris water maze (MWM), are not commercially standardized tests. As a result, there are often deviations from published protocols that can alter the sensitivity of the test. Thus, a method can be referred to as the basis of what was

done, but authors still should describe the apparatus, testing conditions, procedures, and dependent measures that they record and note any changes that might affect the reliability and relevance of the study outcomes. Failure to do so is problematic. Expert guidance on the proper conduct of behavioral studies is available (see, for example, EPA 1998c; Cory-Slechta et al. 2001; Tyl et al. 2008; Makris et al. 2009; Vorhees and Makris 2015; Vorhees and Williams 2015; NAFTA Technical Working Group on Pesticides 2016).

Proper procedural controls and controls for motivation variables are critical for obtaining valid behavioral data. For example, the MWM was used in a number of studies to measure spatial learning and memory, and the most common outcome reported was how long it took an animal to find the hidden platform. To interpret the results, data on swim speed, the use of visible platform control trials, or measures independent of swim speed are needed. Their absence is a serious deficiency in studies that rely on the MWM. Studies included in the monograph that were missing one or more of those controls for the MWM include Zheng et al. (2016), Dong et al. (2017), Zhu et al. (2017), Ge et al. (2018a,b), and Yang et al. (2018). Authoritative reviews are available to help NTP to identify test-specific controls whose absence constitutes a serious deficiency (see, for example, EPA 1998c; Cory-Slechta et al. 2001; Tyl et al. 2008; Makris et al. 2009; Vorhees and Makris 2015; Vorhees and Williams 2015; NAFTA Technical Working Group on Pesticides 2016).

Statistical Analyses

The committee identified several problems related to evaluation of statistical analyses in the animal studies. Although failure to control for litter effects was a critical risk-of-bias factor, the monograph does not appear to give this deficiency proper consideration. For example, Chen et al. (2018), Sudhakar et al. (2018a,b), Sudhakar and Reddy (2018), Sun et al. (2018), Wang et al. (2018), and Zhu et al. (2017) provide no control for litter effects, and these studies were not rated as having a high risk of bias. If NTP acquired information from the study authors, that needs to be clearly documented. Lack of control for litter effects constitutes a critical design and statistical problem in data analyses, and the committee emphasizes the importance of this deficiency for evaluating study validity, particularly in studies that include prenatal exposures.

A second problem is that assumptions made in the appraisals about litters, sample size, and statistics appear to go beyond what the authors of the papers specify. For example, if there were five litters per group in a study, and the results indicate that behavioral outcomes were based on five offspring per group, the monograph seems to indicate that because the sample sizes for litter and offspring are the same, the offspring must have come from five different litters. The committee recommends against such assumptions. Unless stated in the paper or confirmed by NTP via direct communication with the authors, it should never be assumed that the number of offspring tested came from unique litters; experience shows that that is generally not the case. A better assumption is that in the absence

of definitive information, the animals did *not* come from unique litters. Even if offspring came from different litters, if the sex of the offspring is not specified, they could be a mixture of males and females. Males and females exhibit sexually dimorphic behaviors in virtually all behavioral tests, so it is important that results are reported according to sex.

A third problem is that the protocol does not provide sufficient guidance for evaluating statistical analyses. The committee identified a few cases of unacceptable or inappropriate methods, such as the use of multiple t-tests, and inadequate sample sizes. For example, Shalini and Sharma (2015) and Li et al. (2019) used 32 and 45 t-tests, respectively, as the statistical method for identifying group differences, but the Health Assessment Workspace Collaborative (HAWC) appraisal states that the "statistical analyses were reasonable." The use of multiple t-tests that have not been corrected for multiple comparisons or end points has been widely regarded as inappropriate in neurotoxicity research for decades (Muller et al. 1984). A second example is the inappropriate analysis of learning and memory data by using one-way ANOVA. Learning and memory assays require testing over days, usually with multiple trials per day; thus, at a minimum, a repeated-measures ANOVA is required.

Other Potential Threats to Internal Validity

The committee found a threat to internal validity that was not adequately captured in the risk-of-bias criteria described in the protocol. Evidence of severe toxicity as reflected in excessive body-weight loss or significantly lower body-weight gain in exposed compared to control animals is a serious deficiency in developmental-exposure studies. For example, Mesram et al. (2016) reported about 43% lower body weights and about 20% lower brain weights in fluoride-exposed animals compared to controls during early postnatal development. Shalini and Sharma (2015) reported 10% lower body weights and 7% lower brain weights in a 60-day adult fluoride-exposure study. Several adult- and developmental-neurotoxicity studies failed to measure or failed to report maternal, fetal, or pup toxicity (see, for example, Banala et al. 2018; Sudhakar et al. 2018a). Although severe postnatal toxicity was mentioned in risk-of-bias assessments of some studies, it is unclear whether maternal, fetal, or pup toxicity was routinely assessed in all studies. Such effects can seriously confound interpretation of neurodevelopmental effects, and the committee recommends review of those critical variables in neurodevelopmental studies as part of the risk-of-bias assessment.

Exclusion of Studies

Individual studies are normally excluded at the screening stage but can be excluded later in the process if they have a high risk of bias. As noted by NRC (2014, p. 76), "some studies that entail a substantial risk of bias or that have severe methodologic shortcomings ('fatal flaws') [can] be excluded from consideration.

Examples of such exclusion criteria include instability of test compound, inappropriate animal models, inadequate or no controls (or comparison group), or invalid measures of exposure or outcome." NRC (2014) stated that the exclusion criteria should be described in the protocol.

The committee found that some studies cited in the monograph had severe methodologic shortcomings that could potentially warrant exclusion from the body of evidence that informs conclusions about hazard. The shortcomings include evidence of high mortality or severe toxicity, lack of proper controls, and failure to control for litter effects. NTP should define thresholds or conditions for exclusion; authoritative reviews (see, for example, EPA 1998c; Cory-Slechta et al. 2001; Tyl et al. 2008; Makris et al. 2009; Vorhees and Makris 2015; Vorhees and Williams 2015; NAFTA Technical Working Group on Pesticides 2016) can provide a basis for doing so.

Overall Presentation of Study Ratings

The committee had concerns regarding the presentation of risk of bias in the animal evidence. First, the approach to present the evidence differed among the outcomes being considered. For example, risk-of-bias heatmaps were presented separately for lower risk-of-bias and higher risk-of-bias studies of biochemical, neurotransmission, oxidative stress, and histopathology end points (that is similar to what was done with the epidemiology studies), whereas the studies of learning and memory were not stratified in this way. Criteria for stratifying the studies were not presented, so it was unclear how the various risk-of-bias elements were weighted and it was not stated why the learning and memory studies were not stratified. An explicit description of how studies were stratified according to their risk of bias or a ranking of the animal studies is needed for better communication of how the evaluations of individual risk-of-bias elements were integrated to determine the overall quality of any given study. The committee emphasizes that it is not recommending that studies be stratified but to explain clearly what approach is used to evaluate the evidence.

The committee found that the presentation of the risk-of-bias evaluation suffered from a lack of standard ontology for methods and outcomes (see Hardy et al. 2012 and Baker et al. 2018 for relevant ontologic considerations) among the risk-of-bias evaluations and from a the lack of coherence in descriptors between the summary figures of risk of bias presented in the monograph and those in HAWC. For example, the term *open field* as a test method of locomotor activity is used to describe studies that used subjective observer ratings and studies that used automated methods. Those types of studies should not be labeled as equivalent. If such an ontology does not exist, it should be reported as a gap in the report.

DATA EXTRACTION

Several studies (for example, Nageshwar et al. 2017; Nkpaa and Onyeso 2018) involved exposure to fluoride alone and in combination with other treatments, and it was unclear whether NTP evaluated tests of statistical significance appropriately when extracting data from them. That is important because the studies analyzed data collectively, but NTP extracted data only from the control and fluoride-treatment groups, so it was unclear how the effect of fluoride was distinguished from effects in the other treatment groups. A few studies used multiple uncorrected t-tests inappropriately and made all possible comparisons. Did NTP use the t-tests to compare only the control vs the NaF group? How was that done when there were multiple fluoride-dose groups? And what was done when the data were analyzed by ANOVA followed by post hoc individual group comparisons? In those cases, the use of such post hoc tests for a subset of comparisons would not be possible because the error term from the ANOVA includes error variance from all groups, including ones that involved other treatment combinations. A large proportion of the fluoride studies had one to four additional groups exposed to fluoride and a hypothesized "protective" substance; this makes separating fluoride effects from controls difficult or impossible without access to the raw data for reanalyzing the data relevant to the systematic review. The monograph does not describe how that important problem was handled.

CONFIDENCE RATINGS

One consideration in assessing the body of the evidence is sample size, which is critical in neurotoxicity studies.[2] The use of small samples undermines the basis of sampling theory; as sample size decreases, the probability of type II errors increases. For example, the Manusha et al. (2019) study used only five animals per treatment group, and Chen et al. (2018) only six. The evaluation does not mention the very small groups, whose use leads to low study power and difficulties in replication because of large standard errors. The committee was surprised that sample size was apparently not considered as part of NTP's evaluation given that national and international neurotoxicity guidance and guidelines require sample sizes of at least 10 males and females per treatment group (see, for example, EPA 1998a,b) .

[2]Several committee members felt that sample size should be considered as a risk-of-bias element, and individual studies possibly excluded because they used very small samples, but they recognized that that approach is not consistent with current systematic-review methods.

NTP CONCLUSION

On the basis of the information provided in the monograph, it was not obvious to the committee whether NTP followed the protocol to reach its conclusions about the animal evidence. There was no explicit discussion of NTP's confidence in the body of evidence. Rather NTP dismissed all the animal studies by stating that collectively they are "inadequate to inform conclusions on whether fluoride exposure is associated with cognitive effects…in humans" (NTP 2019, p. 2). That conclusion is based on the contention that it is not possible to separate effects on cognitive end points from effects on locomotor activity or motor coordination. In other words, the entire animal dataset on fluoride is essentially dismissed from consideration because changes in locomotor activity in fluoride-exposed animals were regarded as confounding the interpretation of all learning and memory data. The committee questions that rationale. Locomotor-activity changes can sometimes affect learning and memory outcomes, but often they do not, and given that many of the fluoride studies used 3- to 5-min open-field tests that are unreliable measures of locomotor activity, it is inappropriate to use differences in such tenuous assessments to exclude learning and memory results. Moreover, it has been demonstrated many times that the presumed confounding influence of activity on learning and memory behavior does not occur. For example, open-field activity changes do not necessarily translate to swimming tasks. Hence, MWM studies are unlikely to be affected by open-field activity differences unless the activity changes are large and persistent, and this cannot be determined from 3- to 5-min tests of exploration. Even if it could be determined, it must be shown that such locomotor effects are present in a measure, such as swim speed, that directly affects performance in the learning and memory part of the task. If swim speeds are comparable among groups, locomotor-activity differences are not a concern for MWM outcomes. Moreover, swim speed does not affect some dependent measures in the MWM, such as path length and path efficiency. For example, the Yang et al. (2018) study found no effect of fluoride on swimming path length in the MWM—a measure of learning—a result that indicates that any changes found in the 3-min open-field test did not confound findings related to cognition. In the fluoride-neurotoxicity literature, more careful analyses are largely absent, and it is a mistake to dismiss studies of learning and memory because of minor, brief locomotor activity effects or when other assessments can rule out locomotor confounding effects in cognitive assessments.

COMMITTEE CONCLUSIONS

The committee found that some studies had serious deficiencies that made it question the protocol guidance for rating the internal validity of the studies. Several of the risk-of-bias elements appear to need more detail to ensure a rigorous, consistent evaluation of the neurobehavioral studies, and NTP should consider excluding studies that have specific egregious deficiencies (high mortality or severe toxicity, lack of proper controls, and failure to control for litter effects).

The committee also questions NTP's rationale for dismissing the animal evidence as discussed above. Given the serious concerns raised by the committee, NTP will need to decide whether it should reanalyze the animal evidence. The committee cautions, however, that given the poor quality of the animal studies that it reviewed, revising the systematic review to address the concerns highlighted might not affect the finding that the animal evidence is inadequate to inform conclusions about fluoride exposure and neurodevelopmental and cognitive effects in humans.

REFERENCES

Amos-Kroohs, R.M., C.P. Bloor, M.A. Qureshi, C.V. Vorhees, and M.T. Williams. 2015. Effects of developmental exposure to manganese and/or low iron diet: Changes to metal transporters, sucrose preference, elevated zero-maze, open-field, and locomotion in response to fenfluramine,amphetamine, and MK-801. Toxicology Reports 2:1046-1056.

Baker, N., A. Boobis, L. Burgoon, E. Carney, R. Currie, E. Fritsche, T. Knudsen, M. Laffont, A.H. Piersma, A. Poole, S. Schneider, and G. Daston. 2018. Building a developmental toxicity ontology. Birth Defects Res. 110(6):502-518.

Balaji, B., E.P. Kumar, and A. Kumar. 2015. Evaluation of standardized Bacopa monniera extract in sodium fluoride-induced behavioural, biochemical, and histopathological alerations in mice. Toxicol. Ind. Health 31(1):18-30.

Banala, R.R., K.K. Nagapuri, K.P. Mohd, M.M. Reddy, and P.R. Karnati. 2018. Carica papaya leaf extract as a neuroprotective agent against behavioral and neurotransmitter changes in brain of the rat treated with sodium fluoride in pre- and post-natal periods. Pharmacogn. Mag. 14(55):123-131.

Bartos, M., F. Gumilar, C. Bras, C.E. Gallegos, L. Giannuzzi, L.M. Cancela, and A. Minetti. 2015. Neurobehavioural effects of exposure to fluoride in the earliest stages of rat development. Physiol. Behav. 147:205-212.

Chen, J., Q. Niu, T. Xia, G. Zhou, P. Li, Q. Zhao, C. Xu, L. Dong, S. Zhang, and A. Wang. 2018. ERK1/2-mediated disruption of BDNF-TrkB signaling causes synaptic impairment contributing to fluoride-induced developmental neurotoxicity. Toxicology 410:222-230.

Cory-Slechta, D.A., K.M. Crofton, J.A. Foran, J.F. Ross, L.P. Sheets, B. Weiss, and B. Mileson. 2001. Methods to identify and characterize developmental neurotoxicity for human health risk assessment. I: behavioral effects. Environ. Health Perspect. 109(Suppl. 1):79-91.

Curran, C.P., D.W. Nebert, M.B. Genter, K.V. Patel, T.L. Schaefer, M.R. Skelton, M.T. Williams, and C.V. Vorhees. 2011. In utero and lactational exposure to PCBs in mice: Adult offspring show altered learning and memory depending on Cyp1a2 and Ahr genotypes. Environ. Health Perspect. 119(9):1286-1293.

Dong, Y.T., N. Wei, X.L. Qi, X.H. Liu, D. Chen, X.X. Zeng, and Z.Z. Guan. 2017. Attenuating effect of vitamin E on the deficit of learning and memory of rats with chronic fluorosis: The mechanism may involve muscarinic acetylcholine receptors. Fluoride 50(3):354-364.

EPA (US Environmental Protection Agency). 1998a. Health Effects Guidelines: OPPTS 870.6300 Developmental Neurotoxicity Study. Office of Prevention Pesticides and Toxic Substances, US Environmental Protection Agency, Washington, DC. EPA 712-C-98-239.

EPA. 1998b. Health Effects Test Guidelines: OPPTS 870.6200 Neurotoxicity Screening Battery. Office of Prevention Pesticides and Toxic Substances, US Environmental Protection Agency, Washington, DC. EPA 712-C-98-238.

EPA. 1998c. Guidelines for Neurotoxicity Risk Assessment. Risk Assessment Forum, US Environmental Protection Agency, Washington DC. EPA/630/R-95/001F [online]. Available: https://www.epa.gov/sites/production/files/2014-11/documents/neuro_tox.pdf [accessed December 18, 2019].

Ge, Q.D., Y. Tan, Y. Luo, W.J. Wang, H. Zhang, and C. Xie. 2018a. MiR-132, miR-204 and BDNF-TrkB signaling pathway may be involved in spatial learning and memory impairment of the offspring rats caused by fluorine and aluminum exposure during the embryonic stage and into adulthood. Environ. Toxicol. Pharmacol. 63:60-68.

Ge, Y., L. Chen, Z. Yin, X. Song, T. Ruan, L. Hua, J. Liu, J. Wang, and H. Ning. 2018b. Fluoride-induced alterations of synapse-related proteins in the cerebral cortex of ICR offspring mouse brain. Chemosphere 201:874-883.

Hardy, B., G. Apic, P. Carthew, D. Clark, D. Cook, I. Dix, S. Escher, J. Hastings, D.J. Heard, N. Jeliazkova, P. Judson, S. Matis-Mitchell, D. Mitic, G. Myatt, I. Shah, O. Spjuth, O. Tcheremenskaia, L. Toldo, D. Watson, A. White, and C. Yang. 2012. A toxicology ontology roadmap. ALTEX 29(2):129-137.

Kinawy, A.A., and A.A. Al-Eidan. 2018. Impact of prenatal and postnatal treatment of sodium fluoride and aluminum chloride on some hormonal and sensorimotor aspects in rats. Biol. Trace Elem. Res. 186(2):441-448.

Kivrak, Y. 2012. Effects of fluoride on anxiety and depression in mice. Fluoride 45(3 Pt. 2):302-306.

Li, X., J. Zhang, R. Niu, R.K. Manthari, K. Yang, and J. Wang. 2019. Effect of fluoride exposure on anxiety- and depression-like behavior in mouse. Chemosphere 215:454-460.

Makris, S.L., K. Raffaele, S. Allen, W.J. Bowers, U. Hass, E. Alleva, G. Calamandrei, L. Sheets, P. Amcoff, N. Delrue, and K.M. Crofton. 2009. A retrospective performance assessment of the developmental neurotoxicity study in support of OECD test guideline 426. Environ. Health Perspect. 117(1):17-25.

Manusha, S., K. Sudhakar, and K.P. Reddy. 2019. Protective effects of Allium sativum extract against sodium fluoride induced neurotoxicity. Int. J. Pharm. Sci. Res. 10(2):625-633.

Mesram, N., K. Nagapuri, R.R. Banala, C.R. Nalagoni, and P.R. Karnati. 2016. Quercetin treatment against NaF induced oxidative stress related neuronal and learning changes in developing rats. J. King Saud. Univ. Sci. 29(2):221-229.

McPherson, C.A., G. Zhang, R. Gilliam, S.S. Brar, R. Wilson, A. Brix, C. Picut, and G.J. Harry. 2018. An evaluation of neurotoxicity following fluoride exposure from gestational through adult ages in Long-Evans hooded rats. Neurotoxicol. Res. 34(4):781-798.

Muller, K.E., C.N. Barton, and V.A. Benignus. 1984. Recommendations for appropriate statistical practice in toxicologic experiments. Neurotoxicology 5(2):113-125.

NAFTA (North American Free Trade Agreement) Technical Working Group on Pesticides (TWG). 2016. Developmental Neurotoxicity Study (DNT) Guidance Document [online]. Available: https://www.epa.gov/pesticide-science-and-assessing-pesticide-risks/developmental-neurotoxicity-study-guidance [accessed December 18, 2019].

Nageshwar, M., K. Sudhakar, N.C.S. Reddy, and K.P. Reddy. 2017. Neuroprotective effects of curcumin on sodium fluoride induced behavioural and enzymatic changes in brain and muscles of rat. J. Environ. Biol. 38:675-681.

Nkpaa, K.W., and G.I. Onyeso. 2018. Rutin attenuates neurobehavioral deficits, oxidative stress, neuro-inflammation and apoptosis in fluoride treated rats. Neurosci. Lett. 682:92-99.

NRC (National Research Council). 2014. Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, DC: National Academies Press.

NTP (National Toxicology Program). 2016. NTP Research Report: Systematic Review of the Effects of Fluoride on Learning and Memory in Animal Studies. Research Triangle Park, NC: National Toxicology Program. Research Report 1.

NTP. 2017. Protocol for Systematic Review of Effects of Fluoride Exposure on Neurodevelopment. Office of Health Assessment and Translation, Division of the National Toxicology Program, National Institute of Environmental Health Sciences.

NTP. 2019. Draft NTP Monograph on the Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects. Office of Health Assessment and Translation, Division of the National Toxicology Program, National Institute of Environmental Health Sciences, National Institutes of Health, U.S. Department of Health and Human Services.

OECD (Organisation for Economic Co-operation and Development). 2007. Test Guideline 426. OECD Guideline for Testing of Chemicals. Developmental Neurotoxicity Study. Available: https://www.oecd-ilibrary.org/environment/test-no-426-developmental-neurotoxicity-study_9789264067394-en [accessed January 21, 2020].

Pereira, M., P.A. Dombrowski, E.M. Losso, L.R. Chioca, C. Da Cunha, and R. Andreatini. 2011. Memory impairment induced by sodium fluoride is associated with changes in brain monoamine levels. Neurotoxicol. Res. 19(1):55-62.

Raju, S., S.K. Sivanesan, and K. Gudemalla. 2019. Cognitive enhancement effect of Ginkgo biloba extract on memory and learning impairments induced by fluoride neurotoxicity. Int. J. Res. Pharm. Sci. 10(1):129-134.

Sárközi, K., E. Horváth, T. Vezér, A. Papp, and E. Paulik. 2015. Behavioral and general effects of subacute oral arsenic exposure in rats with and without fluoride. Int. J. Environ. Health Res. 25(4):418-431.

Shalini, B., and J.D. Sharma. 2015. Beneficial effects of Emblica officinalis on fluoride-induced toxicity on brain biochemical indexes and learning-memory in rats. Toxicol. Int. 22(1):35-39.

Sharma, C., P. Suhalka, and M. Bhatnagar. 2018. Curcumin and resveratrol rescue cortical-hippocampal system from chronic fluoride-induced neurodegeneration and enhance memory retrieval. Int. J. Neurosci. 129(11):1007-1021.

Sun, Z., Y. Zhang, X. Xue, R. Niu, and J. Wang. 2018. Maternal fluoride exposure during gestation and lactation decreased learning and memory ability, and glutamate receptor mRNA expressions of mouse pups. Hum. Exp. Toxicol. 37(1):87-93.

Sudhakar, K., and K.P. Reddy. 2018. Protective effects of Abelmoschus moschatus seed extract on neurotransmitter system of developing brain of Wistar rats with gestational and post-natal exposure of sodium fluoride. Int. J. Green Pharm. 12(1):S131-S139.

Sudhakar, K., M. Nageshwar, and K.P. Reddy. 2018a. Abelmoschus moschatus extract reverses altered pain and neurohistology of a rat with developmental exposure of fluoride. J. Appl. Pharm. Sci. 8(6):94-104.

Sudhakar, K., M. Nageshwar, and K.P. Reddy. 2018b. Protective effect of okra, Abelmoschus moschatus seed extract on developing brain of rats during pre- and post-natal fluoride exposure. Int. J. Pharm. Sci. Res. 9(4):1519-1528.

Tyl, R.W., K. Crofton, A. Moretto, V. Moser, L.P. Sheets, and T.J. Sobotka. 2008. Identification and interpretation of developmental neurotoxicity effects: a report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints. Neurotoxicol. Teratol. 30(4):349-381.

Vorhees, C.V., and S.L. Makris. 2015. Assessment of learning, memory, and attention in developmental neurotoxicity regulatory studies: synthesis, commentary, and recommendations. Neurotoxicol. Teratol. 52(Pt. A):109-115.

Vorhees, C.V., and M.T. Williams. 2015. Reprint of "Value of water mazes for assessing spatial and egocentric learning and memory in rodent basic research and regulatory studies". Neurotoxicol. Teratol. 52(Pt. A):93-108.

Wang, J., Y. Zhang, Z. Guo, R. Li, X. Xue, Z. Sun, and R. Niu. 2018. Effects of perinatal fluoride exposure on the expressions of miR-124 and miR-132 in hippocampus of mouse pups. Chemosphere 197:117-122.

Yang, L., P. Jin, X. Wang, Q. Zhou, X. Lin, and S. Xi. 2018. Fluoride activates microglia, secretes inflammatory factors and influences synaptic neuron plasticity in the hippocampus of rats. Neurotoxicology 69:108-120.

Zheng, X., Y. Sun, L. Ke, W. Ouyang, and Z. Zhang. 2016. Molecular mechanism of brain impairment caused by drinking-acquired fluorosis and selenium intervention. Environ. Toxicol. Pharmacol. 43:134-139.

Zhu, Y.P., S.H. Xi, M.Y. Li, T.T. Ding, N. Liu, F.Y. Cao, Y. Zeng, X.J. Liu, J.W. Tong, and S.F. Jiang. 2017. Fluoride and arsenic exposure affects spatial memory and activates the ERK/CREB signaling pathway in offspring rats. Neurotoxicology 59:56-64.

4

Human Evidence

The National Toxicology Program (NTP) based its conclusion in the monograph primarily on the human evidence. It considered the human evidence to be "relatively robust" and evaluated the association between fluoride exposure and neurodevelopmental and cognitive effects in 82 publications. It stratified the studies into two categories: lower risk of bias (20 publications[1]) and higher risk of bias (62 publications). Although it evaluated all the studies, the confidence in its conclusion is primarily based on the lower risk-of-bias studies; it concluded that the higher risk-of-bias studies did not affect its confidence in its hazard conclusion. This chapter provides the committee's assessment of NTP's evaluation of the human evidence in the monograph.

LITERATURE SEARCH

In the monograph, NTP clearly displayed the results of the literature search and screening process in a PRISMA flow diagram, a widely accepted framework for reporting a screening process and the ultimate number of included studies. The committee, however, had some concerns regarding NTP's literature-search strategy. One of the sources used to identify articles for the systematic review was the Fluoride Action Network (FAN). The committee acknowledges FAN's efforts in providing several studies that appear to be relevant for the review. However, the process by which FAN identified and selected studies is not clear. FAN identified a number of studies published in Chinese language journals—some of which are not in PubMed or other commonly used databases—and translated them into English. That process might have led to a biased selection of studies and raises the question of whether it is possible that there are a number of other articles in the Chinese literature that FAN did not translate and about which NTP is unaware. NTP should evaluate the potential for any bias that it might have introduced into the literature search process. Possible ways of doing so could include conducting its own searches of the Chinese or other non–English-language literature and conducting subgroup analyses of study quality and results based on the resource used to identify the study (for example, PubMed vs non-PubMed articles). As an initial step in such evaluations, NTP should consider providing empirical information

[1]Two of the 20 publications investigated adults, and the other 18 publications investigated children.

on the pathway by which each of the references was identified. That information would also improve understanding of the sources that NTP used for evidence integration and the conclusions drawn in the monograph. The committee emphasizes that its comments regarding FAN are aimed only at evaluating bias; they are not intended to discourage stakeholder input into the systematic-review process, and the committee acknowledges and encourages the important contributions of FAN and other stakeholder organizations in this process.

STUDY INDEPENDENCE

The unit of analysis in a systematic review is a study, not a report or a publication. The protocol and the monograph do not appear to pay sufficient attention to the independence of multiple publications based on single epidemiologic studies. That is important because NTP enumerates in the monograph and describes in both the text and table summaries specific publications that apparently contribute to the "extent of evidence," one of the criteria on which hazard characterization is based. In at least some cases, a given study is listed and described in separate publications. For example, two publications by Xiang and colleagues (2003, 2011) regarding intelligence (IQ) and fluoride exposures in China were based on the same population, outcome data, and covariates; they are distinguished only by the exposure metric (serum fluoride concentration), which is reported in the 2011 paper to have been collected at the same time as the urine fluoride concentrations in the 2003 paper. By not making it clear that those multiple publications come from a single study, some studies might be double-counted, and NTP's characterization of the extent of the evidence might be exaggerated.

RISK-OF-BIAS EVALUATIONS

Consistency of the Protocol and the Approach

The risk of bias in individual studies was assessed by using criteria provided in the protocol. Three key domains—exposure characterization, outcome assessment, and analysis of potential confounding variables—were emphasized. The committee agrees with the comprehensive approach described but is concerned that there were differences in the approach presented in the protocol compared with that in the monograph. For example, the protocol and the Office of Health Assessment and Translation (OHAT) handbook refer to "Tier 3" studies that are rated as having a high risk of bias in the three key domains (exposure, outcome, and confounding). However, the monograph does not present the studies in tiers but rather categorizes them only as having "higher" or "lower" risk of bias. The approach to assessment of confounding also appears to be somewhat inconsistent. The protocol states that "key covariates" include iodine sufficiency and co-exposure to such neurotoxic compounds as arsenic and lead, and it states that "failure to consider the distribution of the key covariates across the exposure

groups will result in a 'probably high [risk of bias]' or 'definitely high [risk of bias]'" (NTP 2017, p. 9). Those statements in the protocol seem to suggest that the key covariates need to be addressed in every study. However, the monograph states that studies were not required to address every potential confounder but rather only co-exposures or confounders that were considered important for a specific study's population and outcome (NTP 2019, p. 29). Thus, the monograph seems to suggest that the key covariates do not have to be addressed in every study. An example of how that might be a problem is the evaluation of Bashash et al. (2017). NTP does not appear to have considered iodine sufficiency for that study, and arsenic was considered only superficially. However, the study was still rated as "probably low risk of bias" for confounding. That rating might be consistent with the approach described in the monograph, but it seems inconsistent with the protocol. Overall, the protocol and the monograph should be clear and consistent about whether key covariates need to be addressed in every study. If not, the process for deciding which covariates should be addressed in which studies should be clearly described. A final example of inconsistencies between the protocol and the monograph is related to exposure characterization. According to the protocol, "studies that measure or estimate individual exposures, biomarker levels (such as urinary fluoride), or fluoride intake will generally be assigned probably or definitely low [risk of bias] with regard to exposure assessment" (NTP 2017, p. 9). In the monograph, however, Broadbent et al. (2015, p. 73), which used individual "history of use of 0.5-milligram fluoride tablets…and use of fluoridated toothpaste," was rated as having a high risk of bias with respect to exposure assessment.

Thoroughness of the Evaluation

Confounding

NTP developed a reasonable list of the factors most likely to cause confounding in the literature as a whole (NTP 2019, p. 29); in several cases, it provided thoughtful discussions of the likelihood of confounding by some of the factors. For example, NTP identified arsenic as a potential confounder and noted in several cases that studies did not take place in areas known to have high arsenic exposures. The committee, however, identified many cases in which NTP's evaluations or analyses of confounding were insufficient, difficult to understand, or applied inconsistently from one study to another. As noted above, NTP should explain why some sources of potential confounding are considered to be more important in some studies than in others and to address what is known about the magnitude and direction of association between the potential confounders and both fluoride exposure and neurodevelopment. For example, in its analysis of the Russ et al. (2019) study of dementia in the Health Assessment Workspace Collaborative (HAWC), NTP states that "the main confounder missing for evaluating dementia is smoking status." However, the relative risks are low in many studies of smoking and dementia, and this suggests that smoking is unlikely to

have contributed substantially to the hazard ratios of 2.65 (men) and 2.32 (women) of high fluoride exposure and dementia reported in Russ et al. (2019). There are various methods for addressing the potential magnitude of confounding, and NTP should consider some of them (see, for example, Axelson 1978; Rudolph and Stuart 2018).

The potential for confounding might also depend on the source of fluoride. Specifically, the potential confounders that are important in studies where high exposures are due primarily to naturally occurring fluoride in drinking water might differ from those in studies that involve intentionally fluoridated water. For example, arsenic and fluoride may co-occur in some areas that have naturally occurring fluoride, but the co-occurrence might be less common in areas where fluoride exposures come only from intentionally fluoridated water. Overall, the method for assessing which confounders are likely to be important in which studies should include fluoride source.

Exposure Characterization

Exposure misclassification can bias effect estimates in either direction (Jurek et al. 2005). NTP noted in several cases the possibility of a bias from exposure misclassification but did not discuss its likely magnitude and direction and did not discuss it in the context of whether a study reported an association between fluoride exposure and neurodevelopmental and cognitive effects. In many of the studies of childhood neurodevelopment reviewed by NTP, the researchers apparently assessed exposure by using the same methods for all participants regardless of their outcome status. Given that approach, most errors in exposure assessment would most likely bias results to the null. In studies that found no association, that bias could be the major reason for an absence of an association. In studies that identified an association, however, the potential for that bias would be less important in the context of hazard identification because the association would likely be even stronger if one were able to correct for it (Rothman and Greenland 1998).

A possible example can be seen in the Bashash et al. (2017) study of prenatal exposure. Some women had urine samples from all three trimesters, but most did not; this makes the study susceptible to biases resulting from variable completeness of exposure data among participants. Depending on the pattern of "missingness" in relation to true exposure levels (which are unknown) and IQ, the consequences are not readily predictable. Although the bias is not entirely predictable, one might conclude that the missing data could bias results to the null, not toward the association identified in the study. First, there is no indication in the study or other reason to conclude that the number of samples collected from each woman varied strongly by child IQ. Not having samples from all women in all three trimesters would most likely have a nondifferential effect and bias results to the null. Second, the risks might vary by trimester of exposure, and the most susceptible trimester might have been missed in some women. Having exposure data from only a less susceptible trimester in some women would also likely move the results to the null, not toward the positive findings identified.

Another issue related to exposure misclassification can be seen in Broadbent et al. (2015). Here, drinking-water exposures (and thus the differences in exposure) are fairly low. Causal effects are generally more difficult to identify convincingly in studies in which differences in exposure are small. In analyses of fluoride-toothpaste use, participants who reported "always" were compared with those who reported "sometimes." However, if fluoride-toothpaste use is actually relatively high in many people in the "sometimes" category (for example, if they used fluoridated toothpaste 80–99% of the time), the contrast in exposure between the "sometimes" and "always" groups would be small and true effects would be difficult to identify. The same would apply to the comparison of "never" with "ever" use of fluoride pills in the study if many of the participants in the "ever" group used the pills only rarely or if they came from the nonfluoridated parts of the city, which seems likely.

The committee notes that the issues discussed above would probably not change NTP's final risk-of-bias decisions in some cases but might in others. Regardless, failure to address those issues thoroughly and consistently raises the question of whether NTP's evaluations were sufficient and thus whether its final conclusion is based on a fair, transparent, and complete evaluation of the literature.

Outcome Assessment

In assessing cognition or other neurobehavioral outcomes in human studies, it is imperative to protect examiners from information about exposure that could bias their administration and interpretation of assessments. Many neurobehavioral or cognitive assessments require direct interaction with children and interpretation of their responses to test items, so preconceived assumptions about the effects of a specific exposure can result in a biased interpretation in which children assumed to be members of a high-exposure group are classified as more deficient in the outcome. Many of the cross-sectional and case–control studies reviewed by NTP include children from different areas of residence that have different magnitudes of exposure. In those studies, if outcome assessments are conducted in schools or clinics in specific residential areas rather than in a centralized location, children will be identified as belonging to high- or low-exposure groups simply by presenting at those testing locations. Although several studies reviewed by NTP included information on examiner blinding, at least 10 studies did not specify whether outcome assessors were blind to exposure. NTP assumed blinding because urine or drinking-water samples were used to estimate exposure. That assumption can be unfounded, especially in cases in which participants from high- and low-exposure communities were assessed in local schools and clinics where a general sense of exposure characterization could be supposed by the assessor and result in biased outcomes. Because failure to blind examiners can contribute to a high risk of bias in study results and conclusions, this aspect should be considered more carefully in assessing risk of bias in the human studies.

NTP based its conclusions about the effect of fluoride exposure on cognitive and neurodevelopmental outcomes of children on 18 studies that it determined

had lower risk of bias. The studies used a variety of neurodevelopmental and cognitive outcome measures that specifically assessed cognitive development, IQ, attention-deficit hyperactivity disorder (ADHD), visuospatial organization, and memory. Nine of the studies used some form of Raven's Matrices that assesses inductive reasoning by using visual problem-solving tasks. Raven's Matrices do not require verbal responses, so they are often considered the best alternative to standardized intelligence tests based on the English language for assessing cognition in studies of non-English speakers. Use of Raven's Matrices does not increase the risk of bias but assesses a narrow aspect of cognition; it is not equivalent to a full intelligence-test battery that assesses a broad array of cognitive domains. Three of the 18 studies that NTP classified as having low risk of bias used traditional English-based standardized intelligence tests and were accurately classified as having low risk of bias on the basis of the outcome criterion.

In some cases, NTP classified studies as having low risk of bias when the measure of the neurodevelopmental and cognitive outcome was seriously flawed. Given that the outcome determines whether fluoride is hazardous, its proper measurement should be given more weight. One specific example is the study by Barberio (2017) in which the neurodevelopmental outcome is based on parent- or child-reported diagnosis of learning disability or ADHD. That outcome measure is highly problematic because it does not include an objective measure of neurodevelopment or cognition or a confirmation of diagnosis based on review of medical records or objective professional diagnosis. Although NTP recognized that study weakness and judged the outcome as having "probably high risk of bias," the poor quality of the outcome measure warrants a determination of definitely high risk of bias. Furthermore, that weakness should increase the overall risk of bias for the study.

Overall, because of the weaknesses in the tests used in many studies, the committee finds that NTP's assertion (NTP 2019, p. 49) that "it is unlikely that evaluation of additional neurodevelopmental effects would change the hazard conclusion" requires further justification.

Statistical Review

The committee is concerned that the studies included in the systematic review did not undergo a rigorous statistical review. When asked about the role of statisticians in the review process during the committee's public meeting, NTP stated that statisticians were consulted only when the research team was not familiar with the analytic methods used in a study. The committee finds that approach insufficient inasmuch as some of the studies identified as having low risk of bias did not adequately account for the hierarchical structure of their data, and this compromised their internal validity. For example, Ding et al. (2011) sampled children in four elementary schools in China and measured exposure by using urine samples from the children. As demonstrated in Table 1 of that study, water fluoride concentrations differed widely among the communities, and so urine fluoride concentrations were likely highly correlated within the communities. The

study authors, however, failed to account for those relationships, which could have resulted in overly precise interval estimates of the exposure effects and inflated type I error in their statistical tests; this is similar to the effect of ignoring cluster-level treatment assignment in a cluster randomized trial (Cornfield 1978). Similarly, Xiang et al. (2003, 2011) appeared to ignore relationships in exposure between persons from the same village. Unlike the Ding et al. (2011) study, however, proper control for clustering was not possible because there were only two villages. Thus, without control for village effects and given the large differences in fluoride concentrations and IQ between villages, the apparent dose–response relationship could be due to a village effect rather than a fluoride effect. As another example, Green et al. (2019) accounted for community-level effects by adjusting for city in their analysis, but it was unclear how this was done. If they treated city as a random effect, their analytic methods were appropriate. However, if they treated city as a fixed effect, their exposure-effect estimates might be biased. When exposure levels are determined at the group (such as city) level, fixed-effect models do not properly separate exposure effects from group effects, and this results in biased estimates and inflated type I errors (Zucker 1990). Although Green et al. (2019) used individual-level exposure rather than city-level exposure, the fixed-effect model could still produce biased estimates if the exposure levels within a city are highly correlated; this might be expected given that some cities were fully on fluoridated water and others were not. Those analytic issues could have been identified by NTP if statisticians had played a more active role in the development of risk-of-bias instructions or its assessment.

The committee also identified errors in summary statistics that negated the internal validity of some studies that were rated as having low risk of bias. For example, Valdez Jimenez et al. (2017) had multiple errors and internal validity issues among its small cohort of 65 participants. Specifically, there was a large difference in numbers of males and females in the offspring (20 males, 45 females), and apparently incorrect probabilities were reported for age differences between participants and nonparticipants, high rates of cesarean deliveries and premature births among participants (degree of overlap not reported), and incorrect comparisons of observed prematurity rates with national expected rates.

ANALYZING THE DATA

NTP states that all 13 studies of childhood IQ that NTP rated as having a low risk of bias identified at least some evidence of an association of fluoride exposure with neurodevelopmental and cognitive effects (NTP 2019, p. 35). Presented in that way, the numbers suggest a remarkable level of consistency. However, the consistency might be exaggerated if only positive results were selected from studies that reported both positive and negative results (see, for example, Bashash et al. 2017 or Green et al. 2019). At the very least, NTP should acknowledge this issue and provide more context when describing the numbers of positive and negative studies. Alternatively, NTP could develop a series of algorithms a priori that could be used to abstract fully comparable results from the

studies and could then consider the pattern generated by juxtaposing like findings. That approach would avoid selective reporting inasmuch as all studies that generated comparable results would be included. The analysis could be followed by a consideration of the magnitude and consistency of evidence of an association. An example of this type of algorithm can be found in the supplementary material (Web Figure 1) of Carlos-Wallace et al. (2016).

CONFIDENCE RATINGS

Overall

Greater clarity is needed on how the final confidence rating was determined; in some cases, it is not clear whether NTP followed its own procedures. For example, in the monograph (NTP 2019, p. 13) and the protocol (NTP 2017, p. 15), NTP mentions the potential for increasing its confidence in the body of evidence if some criteria, including dose–response relationships and consistency, are present. On the basis of information provided in Figures D1-11 and in several descriptions throughout the monograph, it appears that dose–response patterns were seen in several studies (for example, Xiang et al. 2003; Das and Mondal 2016; Saxena et al. 2012). Furthermore, in a number of sections throughout the monograph, NTP notes the consistency of the evidence. For example, the monograph notes that "all lower risk-of-bias studies in children reported that higher fluoride exposure is associated with at least one measure of decreased IQ or other cognitive effect" (NTP 2019, p. 29). Later in the monograph, NTP states that "the human body of evidence provides a consistent pattern of findings that higher fluoride exposure is associated with decreased IQ in children" (NTP 2019, p. 52). However, despite those statements regarding consistency and the presence of dose–response relationships, NTP does not appear to have increased its confidence ratings for any category of studies (NTP 2019, Table 7). NTP should explain why the confidence ratings did not change for any of the study categories.

Cross-Sectional Studies

NTP's conclusion is based, at least partially, on several cross-sectional studies. Such studies are often criticized because of their potential for reverse causality and exposure misclassification. Reverse causality could involve study subjects in some studies and should be fully evaluated by NTP. However, the committee does not find that reverse causality is likely to be a major concern in most of the cross-sectional studies of fluoride and neurodevelopment identified by NTP because it seems unlikely that diminished neurodevelopmental status would be a widespread and strong determinant of high fluoride exposure in children. Exposure misclassification because of migration in and out of high-fluoride areas could be a concern in some cross-sectional studies but would likely (albeit perhaps not in all cases) bias results toward the null, not toward the positive associations identified

in many studies. In addition, as noted by NTP, several cross-sectional studies minimized such misclassification by including only long-term residents or children who had been living in the same area since birth (see, for example, Xiang et al. 2003, 2011). Overall, the committee felt that well-conducted cross-sectional studies can potentially provide valid and useful information for evaluating the effects of fluoride on neurodevelopment and thus agrees with NTP that these studies should not necessarily be given a final rating of "lower confidence." As an aside, the committee did not agree with NTP's use of the term *functionally prospective* to describe some cross-sectional studies. NTP did not define or explain that term, and it is not used in the epidemiology literature; therefore, the committee discourages its use.

Publication Bias

NTP seems to be relying on the results of the funnel plot from the meta-analysis of Choi et al. (2012) for its analysis of publication bias. Although the lack of asymmetry in the plot provides some evidence against major publication bias, NTP should acknowledge the weaknesses of the approach; for example, factors other than publication bias can affect the symmetry of funnel plots, and funnel plots rely on subjective interpretation. In addition, the monograph includes a number of studies that were not included in the Choi et al. (2012) meta-analysis. Thus, NTP should do its own analyses of publication bias and use the analyses to evaluate the likelihood that publication bias could have had major effects on the body of evidence that it has identified.

SUMMARIZING AND PRESENTING THE DATA

Definitions of *Consistency* and *Positive*

A key conclusion of the monograph is that the results of the epidemiologic studies consistently show a positive association. Although the desire to provide a simple summary of a complex array of evidence is understandable, such claims imply that the studies provide an array of clearly comparable results and that all suggest an adverse effect of fluoride on neurodevelopment or cognition. In fact, many of the studies provide results that are based on multiple indicators of fluoride exposure, assess multiple measures of cognition and neurodevelopment at different ages, use multiple statistical approaches to characterize the relationship between fluoride exposure and health outcomes, and address markedly different magnitudes of fluoride exposure. The committee recognizes that drawing conclusions always requires aggregating or summarizing data that have some degree of heterogeneity, but the data should be examined as subsets along one or more of the axes suggested above. For example, what do the studies of urinary fluoride resulting from naturally occurring fluoride exposure indicate for IQ below the age of 5 years? Accordingly, the monograph should juxtapose results of broadly comparable studies and use the resulting information to provide a text

summary of the patterns observed. If comparing "like to like" results yielded consistent results for all measures, ages, exposure sources, statistical approaches, and exposure ranges—taking random error into account—that would indeed warrant a statement that results consistently show adverse effects. The monograph, however, does not provide the evidence in a manner that leads to that conclusion.

The text that is used to justify the assertions of consistently positive results is purely anecdotal and cites isolated findings from specific studies without explaining why those findings, and not others, were highlighted. Selective reporting of the literature in that way is almost certain to generate a false impression of consistency. Although it might be true that every study has at least *some* indication of adverse outcomes associated with higher fluoride exposure, that does not provide a clear or necessarily useful assessment of the body of evidence. Furthermore, it is inappropriate to rely on statistical significance as the single indicator of whether a study is called "positive," given that studies with low power can nonetheless generate an indication of a positive association and that those with isolated statistically significant findings might not provide an overall pattern indicative of a positive association. The information provided in the monograph does not allow readers to follow the steps from assembling and presenting available data in an objective and informative manner to making observations about the pattern of the results to drawing conclusions from the patterns that were observed.

Methodical Presentation of Results

A full understanding of the data calls for their detailed examination in relation to the methodologic features of the studies. Study results can be arrayed in multiple ways that would be informative. For example, studies can be categorized on the basis of such risk-of-bias criteria as blinding or such factors as the major source of fluoride in each study (naturally occurring vs intentional addition), magnitudes of exposure, or the ages at which an exposure or outcome was assessed. Informative evaluations can then be made by comparing study results within the categories. For example, if the methodologically strongest studies tend to show clearer associations than the methodologically weaker ones, the evidence could be interpreted as providing greater support for a possible adverse effect than if the reverse were found. Consistency among studies of varied methodologic quality might also help to provide evidence that some issues do not present major concerns. For example, if studies in which the researchers were blinded yield results similar to those of a comparable set of studies in which researchers were not blinded, this finding might provide evidence that failure to blind was not a major source of bias. Similarly, categorizing studies on the basis of exposure might help to identify dose–response relationships; that is, if a true association exists, studies that have the highest exposures and the widest range between the "low" and "high" exposure groups would be expected to report greater effect sizes than studies that involve low exposures and a narrow exposure range between groups. Overall, by categorizing studies on the basis of a variety of methodologic factors

and comparing groups of studies within the different categories, NTP should be able to provide a more detailed and comprehensive evaluation of the literature.

The committee notes that the presentation of study results in Table 6 and Figures D1-12 made it difficult to assess the variability of exposure–effect estimates across studies. It recommends that NTP present forest plots—similar to Figure 2 in Choi et al. (2012)—for subgroups of studies that have the same effect estimate (such as relative risk), that have similar dose and outcome measurements, and that adjust for the same set of confounders. The committee also notes that the numbers of studies in particular categories, such as those grouped by study design, are inconsistently stated throughout the monograph; these inconsistencies should be corrected.

Finally, the committee agrees with NTP's decision to base its conclusions primarily on studies that have a lower risk of bias given the previous discussion regarding NTP's risk-of-bias evaluations. However, its focus on the lower risk-of-bias studies of childhood neurodevelopment outcomes and what seem to be the highly consistent findings across all these studies might give the impression that NTP has artificially increased the confidence in its conclusion regarding this outcome. Stratifying the higher risk-of-bias studies (NTP 2019, Figure A3-1) vs lower risk-of-bias studies (NTP 2019, Figure A3-3) into separate figures might be one source of that concern. NTP should consider creating one figure that includes the risk-of-bias ratings for all the studies with a stratification that separates the categorized higher risk-of-bias vs lower risk-of-bias studies. That approach would represent better how it considered the body of literature by assessing all studies but focusing its conclusions on the lower risk-of-bias studies.

Rationale for Not Performing a Meta-Analysis

The committee strongly recommends that NTP reconsider its decision not to perform a meta-analysis and, if it still decides not to do a meta-analysis, that it provide a more thorough and convincing justification for its decision. In the monograph, NTP states that a meta-analysis was not performed because of "heterogeneity in dose among the available human evidence, and because a hazard conclusion could be reached without conducting a meta-analysis" (NTP 2019, p. 13). A properly conducted meta-analysis can account for heterogeneity in exposure measurements and other aspects of study design, so it is not clear why heterogeneity was listed as a reason for not performing one. It would be difficult to perform one meta-analysis that includes both relative risk estimates and mean differences (or standardized mean differences), but these could be separated out into two meta-analyses. Potentially, meta-analyses of studies deemed sufficiently similar in their exposure and outcome metrics could also be performed and could address NTP's concern about heterogeneity. However, because NTP did not present the studies in a way that would suggest such groupings, the committee is unclear how feasible such analyses would be. The committee also recommends that NTP explain why it did not update the Choi et al. (2012) meta-analysis. NTP uses the funnel plot in Choi et al. (2012) as evidence of minimal publication bias in its

systematic review. However, Choi et al. (2012) considered only a subset of the studies included in the systematic review, so NTP's claim of minimal publication bias would be strengthened by adding recent papers to the meta-analysis and constructing a new funnel plot.

Communication Regarding Lower Exposures

The discussion section of the monograph provides an informal assessment of the evidence with regard to exposure range and declares that the positive results are based largely on exposures greater than those used for fluoridation. The basis of that inference is not apparent, and it seems to contradict the earlier assertion that nearly all the studies are positive, including ones that assessed lower exposures. More important, as discussed in Chapter 2, this discussion gives a false impression that NTP conducted a formal dose–response assessment. NTP should make it clear that the monograph cannot be used to assess what concentrations of fluoride are safe.

NTP CONCLUSION

The monograph "concludes that fluoride is presumed to be a cognitive neurodevelopmental hazard to humans. This conclusion is based on a consistent pattern of findings in human studies across several different populations showing that higher fluoride exposure is associated with decreased IQ or other cognitive impairments in children" (NTP 2019, p. 59). The committee was tasked with assessing whether NTP satisfactorily supports its conclusion. In light of the issues raised by the committee regarding the analysis of various aspects of some studies and the analysis, summary, and presentation of the data in the monograph, the committee finds that NTP has not adequately supported its conclusion. The committee's finding does not mean that the conclusion is incorrect; rather, further analysis or reanalysis as suggested in the present report is needed to support the conclusion in the monograph.

REFERENCES

Axelson, O. 1978. Aspects on confounding in occupational health epidemiology. Scand. J. Work Environ. Health 4:85-89.

Barberio, A.M., C. Quinonez, F.S. Hosein, and L. McLaren. 2017. Fluoride exposure and reported learning disability among Canadian children: Implications for community water fluoridation. Can. J. Public Health 108:229-239.

Bashash, M., D. Thomas, H. Hu, E.A. Martinez-Mier, B.N. Sanchez, N. Basu, K.E. Peterson, A.S. Ettinger, R. Wright, Z. Zhang, Y. Liu, L. Schnaas, A. Mercado-Garcia, M.M. Tellez-Rojo, and M. Hernandez-Avila. 2017. Prenatal fluoride exposure and cognitive outcomes in children at 4 and 6-12 years of age in Mexico. Environ. Health Perspect. 125(9):1-12.

Broadbent, J.M., W.M. Thomson, T.E. Moffitt, and R. Poulton. 2015. Community water fluoridation and intelligence response. Am. J. Public Health 105:3-4.

Carlos-Wallace, F.M., L. Zhang, M.T. Smith, G. Rader, and C. Steinmaus. 2016. Parental, In Utero, and Early-Life Exposure to Benzene and the Risk of Childhood Leukemia: A Meta-Analysis. Am. J. Epidemiol. 183(1):1-14.

Choi, A.L., G. Sun, Y. Zhang, and P. Grandjean. 2012. Developmental fluoride neurotoxicity: A systematic review and meta-analysis. Environ. Health Perspect. 120:1362-1368.

Cornfield, J. 1978. Randomization by group: a formal analysis. Am. J. Epidemiol. 108(2): 100–102.

Das, K., and N.K. Mondal. 2016. Dental fluorosis and urinary fluoride concentration as a reflection of fluoride exposure and its impact on IQ level and BMI of children of Laxmisagar, Simlapal Block of Bankura District, W.B., India. Environ. Monit. Assess. 188:218.

Ding Y., H. Sun, H. Han, W. Wang, X. Ji, X. Liu, and D. Sun. 2011. The relationships between low levels of urine fluoride on children's intelligence, dental fluorosis in endemic fluorosis areas in Hulunbuir, Inner Mongolia, China. J. Hazard Mater. 186:1942-1946.

Green R., B. Lanphear, R. Hornung, D. Flora, E.A. Martinez-Mier, R. Neufeld, P. Ayotte, G. Muckle, and C. Till. 2019. Association between maternal fluoride exposure during pregnancy and IQ scores in offspring in Canada. JAMA Pediatr. E1-E9.

Jurek, A.M., S. Greenland, G. Maldonado, and T.R. Church. 2005. Proper interpretation of non-differential misclassification effects: expectations vs observations. International Journal of Epidemiology 34:680–687.

NTP (National Toxicology Program). 2017. Protocol for Systematic Review of Effects of Fluoride Exposure on Neurodevelopment. Office of Health Assessment and Translation, Division of the National Toxicology Program, National Institute of Environmental Health Sciences.

NTP. 2019. Draft NTP Monograph on the Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects. Office of Health Assessment and Translation, Division of the NTP, National Institute of Environmental Health Sciences, National Institutes of Health, US Department of Health and Human Services.

Rothman, K.J., and S. Greenland. 1998. Modern Epidemiology. 2nd edn. Lippincott-Raven, U.S.A.

Rudolph, K.E., and E. A. Stuart. 2018. Using sensitivity analyses for unobserved confounding to address covariate measurement error in propensity score methods. Am. J. Epidemiol. 187(3):604-13.

Russ, T.C., L.O.J. Killin, J. Hannah, G.D. Batty, I.J. Deary, and J.M. Starr. 2019. Aluminum and fluoride in drinking water in relation to later dementia risk. Brit. J. Psychol. 14:1-6.

Saxena, S., A. Sahay, and P. Goel. 2012. Effect of fluoride exposure on the intelligence of school children in Madhya Pradesh, India. J. Neurosci. Rural Pract. 3:144-149.

Valdez Jimenez, L., O.D. Lopez Guzman, M. Cervantes Flores, R. Costilla-Salazar, J. Calderon Hernandez, Y. Alcaraz Contreras, and D.O. Rocha-Amador. 2017. In utero exposure to fluoride and cognitive development delay in infants. Neurotox. 59:65-70.

Xiang, Q., Y. Liang, L. Chen, C. Wang, B. Chen, X. Chen, and M. Zhou. 2003. Effect of fluoride in drinking water on children's intelligence. Fluoride 36:84-94.

Xiang, Q., Y. Liang, B. Chen, and L. Chen L. 2011. Analysis of children's serum fluoride levels in relation to intelligence scores in a high and low fluoride water village in China. Fluoride 44:191-194.

Zucker, D. M. 1990. An analysis of variance pitfall: The fixed effects analysis in a nested design. Educational and Psychological Measurement. 50(4):731-738.

Appendix

Biographic Information on the Committee to Review the NTP Monograph on the Systematic Review of Fluoride Exposure and Neurodevelopmental and Cognitive Health Effects

David A. Savitz (Chair) is professor of epidemiology and associate dean for research of the Brown University School of Public Health, with joint appointments in obstetrics and gynecology and pediatrics at the Alpert Medical School. He was vice president of research at the university from 2013 to 2017. His epidemiologic research has addressed a wide array of important public-health issues, including environmental hazards in the workplace and community, reproductive health outcomes, and environmental influences on cancer. He has worked extensively on health effects of nonionizing radiation, pesticides, drinking-water treatment byproducts, and perfluorinated compounds. Before joining Brown University, Dr. Savitz held appointments as the Charles W. Bluhdorn Professor of Community and Preventive Medicine at Mount Sinai School of Medicine and professor at the University of North Carolina School of Public Health. He was president of the Society for Epidemiologic Research and the Society for Pediatric and Perinatal Epidemiologic Research and was a North American regional councilor for the International Epidemiological Association. Dr. Savitz was elected to the National Academy of Medicine in 2007. He received an MS in preventive medicine from Ohio State University and a PhD in epidemiology from the University of Pittsburgh Graduate School of Public Health.

Germaine M. Buck Louis is dean of the College of Health and Human Services of George Mason University. Her research has addressed a mixture of environmental exposures, including endocrine disruptors, stress, diet, and physical activity in relation to a spectrum of reproductive outcomes in men and women. She was an early pioneer in the application of the exposome research paradigm for understanding environmental influences on human fecundity and fertility impairments. Before joining the university, Dr. Louis was the director of the Division of Intramural Population Health Research in the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the Na-

tional Institutes of Health, where she led population-health scientists in designing research aimed at enhancing the health and well-being of fetuses, pregnant women, children, and young adults. She has served the National Academies, Pan American Health Organization, US Environmental Protection Agency, and World Health Organization in various roles. She is a former president of the Society of Pediatric and Perinatal Epidemiologic Research and of the Society for Epidemiologic Research and has served on the boards of the American College of Epidemiology and the International Society for Environmental Epidemiology. Dr. Louis received a PhD in epidemiology from the State University of New York at Buffalo.

Kevin M. Crofton is principal and consultant at R3Fellows, LLC. Previously, he worked for more than 35 years as a developmental neurotoxicologist in the US Environmental Protection Agency (EPA) Office of Research and Development. Dr. Crofton has also served as an adjunct associate professor at Duke University, the University of North Carolina, and North Carolina State University. His research interests include developmental neurotoxicity with an emphasis on understanding the consequences of endocrine disruption on neurodevelopment. He recently received the EPA Distinguished Career Service Award. Dr. Crofton received an MS in toxicology from Miami University and a PhD in toxicology from the University of North Carolina at Chapel Hill.

Akhgar Ghassabian is an investigator and assistant professor in the Departments of Pediatrics, Population Health, and Environmental Medicine of the New York University (NYU) School of Medicine. Her research focuses on identifying environmental exposures that contribute to the etiology of developmental disabilities in childhood. Before joining NYU, Dr. Ghassabian was the intramural research training award fellow at the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health. During her doctoral and postdoctoral training, Dr. Ghassabian was involved in birth-cohort studies in Europe and in the United States. She was a collaborator on European epidemiologic consortia examining the effect of nutrition and air pollution on children's neurodevelopment. Dr. Ghassabian was the recipient of the Rubicon Award from the Netherlands Organization for Scientific Research in 2014 and the Robin/Guze Young Investigator Award from the American Psychopathological Association in 2019. She obtained an MD from Tehran University of Medical Sciences and a PhD in epidemiology from Erasmus University Rotterdam, the Netherlands.

Judith B. Klotz is an affiliate faculty member in the Department of Environmental and Occupational Health of the Dornsife School of Public Health, Drexel University, and an adjunct associate professor in the Department of Epidemiology of the Rutgers School of Public Health. She is a member of the Health Effects Committee of the New Jersey Drinking Water Quality Institute and of the Public Health standing committee of the Science Advisory Board, both

advisory groups of the New Jersey Department of Environmental Protection. She served as environmental scientist and program manager in environmental health and in cancer surveillance in the New Jersey Department of Health from 1984 to 2003 and focused especially on toxic substances in drinking water and environmental epidemiology of cancer and reproductive outcomes. Dr. Klotz has served on several National Academies committees, including the Committee on Fluoride in Drinking Water and the Committee on the Review of the Styrene Assessment in the National Toxicology Program 12th Report on Carcinogens. She received an MS in genetics from the University of Michigan and a DrPH in environmental health sciences from Columbia University.

Juleen Lam is an assistant professor in the Department of Health Sciences of the California State University, East Bay. She is also an affiliate researcher in the Department of Obstetrics, Gynecology and Reproductive Sciences of the University of California, San Francisco, School of Medicine. Her research interests are in environmental epidemiology, evaluation of population exposures to environmental contaminants, assessment and communication of environmental risks, and reproductive and developmental health. She specializes in analysis of environmental-health data and development and application of risk-assessment methods. Dr. Lam has been involved in the development of systematic review methods for environmental-health data and has had a pivotal role in implementing, publishing, and disseminating these approaches in academic and government settings. She is a member of the US Environmental Protection Agency Board of Scientific Counselors Chemical Safety for Sustainability Subcommittee. She is currently serving on the National Academies Committee to Review DOD's Approach to Deriving an Occupational Exposure Limit for TCE. She received an MS in environmental engineering management from George Washington University and an MHS in biostatistics and PhD in environmental-health policy from the Johns Hopkins University Bloomberg School of Public Health.

Pamela J. Lein is a professor of neurotoxicology in the Department of Molecular Biosciences of the University of California, Davis, School of Veterinary Medicine. Her research interests are in how environmental stressors interact with genetic susceptibilities to influence the risk and severity of neurodevelopmental disorders and neurodegeneration. Because altered patterns of connectivity are associated with neurologic deficits, her research focuses on investigating how environmental contaminants, chemical convulsants, and inflammation perturb neuronal connectivity as determined by using biochemical, morphogenic, and electrophysiologic end points. Her group is also developing biomarkers of organophosphate neurotoxicity and testing novel therapeutic approaches for protecting against the neurodegenerative effects associated with neurotoxic proconvulsants. Dr. Lein was a member of the National Academies Committee to Review Report on Long-Term Health Effects on Army Test Subjects. She

received an MS in environmental health from East Tennessee State University and a PhD in pharmacology and toxicology from the State University of New York at Buffalo.

Michael L. Pennell is associate professor in the Division of Biostatistics in the College of Public Health of Ohio State University. His research interests are in nonparametric Bayes, first hitting time models for survival analysis; design and analysis of group randomized trials; joint modeling outcomes of different scales; statistical methods in toxicologic risk assessment; and statistical applications in biomedical research, including cancer control, pathology, and veterinary medicine. Dr. Pennell has served as an ad hoc member of the US Environmental Protection Agency (EPA) Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel, the EPA Science Advisory Board on trichloroethylene and Libby amphibole asbestos, and the Chemical Safety Advisory Subcommittee for 1-bromopropane. He currently serves on the National Academies Committee to Evaluate the IRIS Protocol for Inorganic Arsenic. He received an MS and a PhD in biostatistics from the University of North Carolina at Chapel Hill.

Craig Steinmaus is an associate adjunct professor of epidemiology at the University of California, Berkeley (UCB). He is also a public-health medical officer III in the California Environmental Protection Agency (CalEPA) and is the UCB director of the Arsenic Health Effects Research Group. He is a board-certified physician with over 12 years of patient-care experience. His epidemiologic research has involved studies of drinking-water contaminants with a focus on early-life exposure and other factors conferring susceptibility. He also teaches graduate courses on epidemiology, causal inference, and systematic review at UCB and at the University of California, San Francisco. Dr. Steinmaus has served on several study sections of the National Institutes of Health and Centers for Disease Control and Prevention and is a full member of the Cancer, Heart, and Sleep Epidemiology, A study section. His work in the CalEPA water toxicology section has involved systematic reviews and risk assessments of drinking-water agents, including nitrate, arsenic, copper, perchlorate, fluoride, chromium, and trihalomethanes. He received an MD from the University of California, Davis School of Medicine and an MPH in environmental-health sciences from UCB.

Charles V. Vorhees is a professor in the University of Cincinnati College of Medicine. He is co-director of the Animal Behavior Core and program director of the Teratology Training Program. He is on the graduate faculty of the Graduate Programs in Neuroscience and Molecular and Developmental Biology. His research focuses on brain development and behavior. He was a founding member of the Neurobehavioral Teratology Society in 1977 and was elected president in 1984–1985 and 2012–2013. Dr. Vorhees has served on multiple scientific advisory committees for the US Food and Drug Administration, US

Environmental Protection Agency, and National Institutes of Health. He was on the National Academies Subcommittee on Reproductive and Developmental Toxicants. Dr. Vorhees obtained an MA and a PhD in neurobiology from Vanderbilt University.

Kimberly Yolton is a professor in Cincinnati Children's Hospital Medical Center (CCHMC) and the University of Cincinnati College of Medicine and director of research in the Department of General and Community Pediatrics. She is a developmental psychologist and epidemiologist with over 25 years of experience studying the effects of prenatal and early-life exposures on neurobehavior from infancy through childhood and directs the longitudinal Health Outcomes and Measures of the Environment (HOME) Study. She was formerly the director of a follow-up clinic serving high-risk infants and young children and has extensive experience with infants and children who were prenatally exposed to substances of abuse, who were born prematurely or at low birth weight, or who come from disadvantaged home environments. She was involved in the initial development of the NICU Network Neurobehavioral Scale (NNNS), a specialized neurobehavioral assessment tool used with healthy and high-risk newborns, and conducts frequent training on the proper administration, scoring, and interpretation of the instrument for research and clinical purposes. She has been affiliated with the National Institutes of Health–funded Neonatal Research Network for over 25 years at two sites as an examiner, Gold Standard reviewer for intelligence testing, follow-up principal investigator, and steering-committee member. She often collaborates with investigators regarding neurobehavioral assessment and staff training strategies to acquire the most appropriate outcome measures with the highest standards of reliability and validity. She earned a PhD in child development and developmental psychology from Ohio State University and completed a 3-year National Research Service Award in Pediatric Environmental Health at CCHMC.

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA

Case Number: 3:17-cv-02162-EMC

PLTF / DEFT EXHIBIT
NO. 653

Date
Admitted:_____

By:_____

Vicky Ayala, Courtroom Deputy
